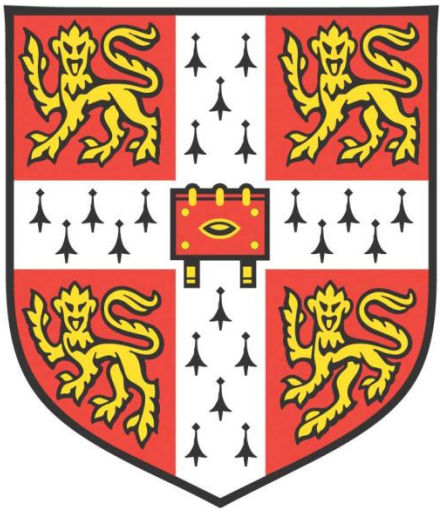
TOWARDS UNDERSTANDING HELICASE AND CHAPERONE ACTIVITIES IN THE RNA DEGRADOSOME



Zi Ran Shen

Wolfson College

Department of Biochemistry

University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy

January 2020

DECLARATION

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. This thesis is does not exceed 60,000 words excluding tables, footnotes, bibliography, and appendices.

ABSTRACT

The *E. coli* RNA degradosome is a complex multi-enzyme machine that is central to the post-transcriptional regulation of the cell. Some of its functions include maturing and processing sRNA, rRNA, and tRNA, as well as degrading mRNA. The key components of the RNA degradosome include the endoribonuclease RNase E, the DEAD-box RNA helicase RhlB, the glycolytic enzyme enolase, and the phosphorolytic exoribonuclease PNPase. The degradosome has also been previously shown to associate with the RNA chaperone Hfq to form a small RNA guided machinery that targets defined transcripts. This thesis attempts to investigate several characteristics of the degradosome, including the importance of RhlB, the structure of a portion of the RNase E C-terminal domain that recruits enolase and helicase, and the association between Hfq, ChiX and the C-terminus of RNase E. The thesis also explores structural details regarding the small domain of RNase E involved in both RNA binding and oligomerisation and its relationship to the RNA-binding KH domains. Utilizing a point mutation in the DEAD box of RhlB, I have found increased RNA affinity to mutant RhlB, the potential structure of Hfq — an RNA chaperone — bound to the sRNA ChiX, and a structural correlation between the RNase E small domain and other KH domains. Preliminary models of Hfq and ChiX structure show a novel binding mode for class II sRNAs as the majority of ChiX associates with the distal face of Hfq. Bioinformatic studies reveal evolutionary roots between the KH domains and the RNase E small domain, supporting the hypothesis that the RNase E small domain may be involved in a novel mode of RNA binding and recognition.

ACKNOWLEDGEMENTS

I would like to thank many people who have helped me in my PhD — if I were to list them all this PhD will surely exceed 60,000 words in names alone. In the interest of brevity, I will list only those integral to my life in Cambridge but rest assured I am deeply grateful for everyone who has helped me along the way, however briefly. I'd like to first and foremost thank Professor Ben Luisi, one of the kindest and most supportive people I've ever met in my life. Without him I would've never set foot in Cambridge, let alone finish this thesis. I'd also like to thank my lab mate Tom Dendooven for his endless advice on cryo-EM for which I am forever grateful and perpetually in awe. I'd also like to thank my lab mates (special shoutout to Miao, Kotryna, and Heather, and Tai) for all of your friendship and support. Thank you kindly to my friends at Wolfson College who brought so much joy to my live over the last four years. A big thank you to my parents, who have selflessly provided me with love and education throughout my life. Last but certainly not least, I'd like to thank my partner Bas Monsewije for all the highlights of my Cambridge experience — he continues to make me happier even though I am the happiest I've ever been. Thank you to everyone who's been a part of this amazing journey with me.

CONTENTS

List of Tables

LIST OF FIGURES

LIST OF ABBREVIATIONS

AR1 ..... Arginine Rich region 1
AR2 ..... Arginine Rich region 2
cryoEM ..... Cryo-Electron Microscopy
CTF ..... Contrast Transfer Function
CV ..... Column Volume
ddH_2O ..... Double Distilled Dater, or Distilled Deionized Water
EBS ..... Enolase Binding Site
EMSA ..... Electrophoretic Mobility Shift Assay
GS-FSC ..... Gold Standard Fourier Shell Correlation

HIC ..... Hydrophobic Interaction Column
IPTG ..... Isopropyl β- d-1-thiogalactopyranoside
MTS ..... Membrane Targeting Sequence
PALM ..... Photo-activated Localization Microscopy
PBS ..... PNPase Binding Site
PCR ..... Polymerase Chain Reaction
PNPase ..... Polynucleotide Phosphorylase
RBS ..... RhlB Binding Site
RhlB ..... RNA helicase B
SD ..... Shine Dalgarno
SDS-PAGE ..... Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
sRNA ..... small RNA
STORM ..... Stochastic Optical Reconstruction Microscopy
STED ..... Stimulated Emission Depletion Microscopy
TBE ..... Tris/Borate/EDTA
Ternary complex ..... RNase E 603-850, RhlB, Enolase
UTR ..... Untranslated Region

1 INTRODUCTION

1.1 Overview

The prokaryotic organism is constantly facing new challenges and adversities in its environment. To survive, it must be able to adapt rapidly to each new environmental input, without compromising essential cellular functions. When nutrients are abundant, they are rapidly imported and used for cell growth and division. When conditions become astringent, conservational methods must be employed. Maintaining a dynamic equilibrium between growth and conservation allows for cells to be flexible in response to changes in the environment. Expeditious actions in the cell can best occur via post-transcriptional regulation, as regulating gene expression at the transcriptional level requires an abundance of both time and energy, resources which are both scarce and valuable. RNA levels, in contrast, can change in minutes. Changes in RNA levels can rapidly affect protein levels in the cell without being as expensive energetically. As post-transcriptional regulation plays an integral role in prokaryotic survival, the understanding of this process is thus vital to the understanding of prokaryotic life.

In the mid-80's, a paper was published detailing a novel method of post-transcriptional regulation by a small RNA, MicF (Coleman et al, 1984). MicF was shown to regulate *ompF*, a gene encoding an outer membrane protein (Aiba et al, 1987). Interestingly, the gene of *micF* has its own promoter and is independently transcribed, though it only encodes a small RNA of 93 nt, a phenomenon which was unheard of at the time (Andersen et al, 1987). Not only was MicF independently transcribed, it was also found to be active by exerting a post-transcriptional regulatory effect by repressing *ompF* translation (Aiba et al, 1987). When MicF was bound to

the Shine-Dalgarno sequence of *ompF*, *ompF* translation was suppressed (Andersen & Delihas, 1990). Once suppressed, MicF was found to facilitate the degradation of the *ompF* transcript as well (Figure 1) (Andersen & Delihas, 1990). Since this discovery, many small RNAs, hereby denoted as sRNAs, have been found in almost all bacterial families. Discovered sRNAs have not only shown a surprising amount of variation in length but also in mechanism of action, prompting further research into each individual sRNA (Wagner & Romby, 2015).

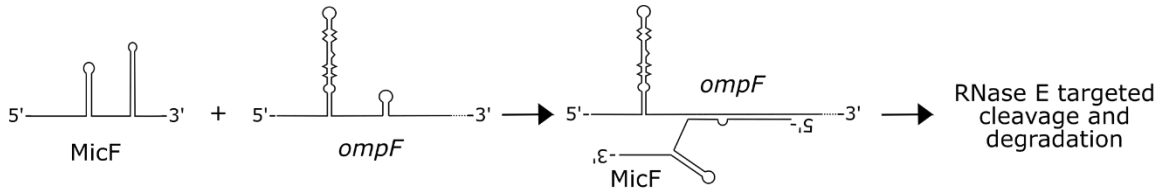


Figure 1: Mechanism of *ompF* repression by MicF. In the environment of an abundance of nutrients, MicF is actively transcribed in the cell. Increased concentrations of cellular MicF lead to base-pairing with *ompF* at both the Shine-Dalgarno box and translation initiation site. Translation of *ompF* is thus unable to occur, leaving *ompF* vulnerable to RNase E mediated cleavage and decay. In the absence of nutrients, MicF is bound by StpA which targets it for degradation, leaving the 5'UTR of *ompF* free for translation initiation.

Around the time of MicF's discovery, the ribonuclease RNase E and its various RNA processing functions were also beginning to emerge. When RNase E genes, then known as *ams*, *rne*, or *hmp*, were mutated, ribosomal RNAs and RNA decay were found to be disrupted (Cohen & McDowall, 1997). The protein RNase E was then identified as a key enzyme in RNA metabolism. Since then, RNase E has undergone scrutiny in various ways, increasing the pool of knowledge regarding RNA metabolism since the 80's. We now know that RNase E is the scaffold upon which the multi-enzyme machine known as the RNA degradosome assembles,

as well as being an endonucleolytic enzyme responsible for tRNA, rRNA, and endogenous sRNA maturation (Bandyra & Luisi, 2018). The RNA degradosome is a multienzyme complex consisting of RNase E, RNA helicase B (RhlB), enolase, and polynucleotide phosphorylase (PNPase). These enzymes work together to ensure the correct RNA is degraded or processed in the prescribed manner. The RNA degradosome, aside from maturing RNA, is also implicated in sRNA mediated post-transcriptional regulation (Bandyra & Luisi, 2018). As such, the RNA degradosome is an indispensable element of RNA regulation and control. In addition to the RNA degradosome, other proteins such as Hfq have also been shown to participate in sRNA mediated post-transcriptional regulation. Hfq is an RNA chaperone that facilitates sRNA binding to its target mRNA and sometimes the recruitment of the degradosome as well (Bruce et al, 2018). Though the method by which Hfq recruits the degradosome remains unclear, the importance of Hfq's role in RNA mediation is irrefutable, as ΔHfq strains have been shown to have decreased virulence and heightened sensitivity to stress conditions (Chao & Vogel, 2010). Together, Hfq and the RNA degradosome modulate the concentrations of RNA in the cell, and thus mediate the rate of translation and protein synthesis.

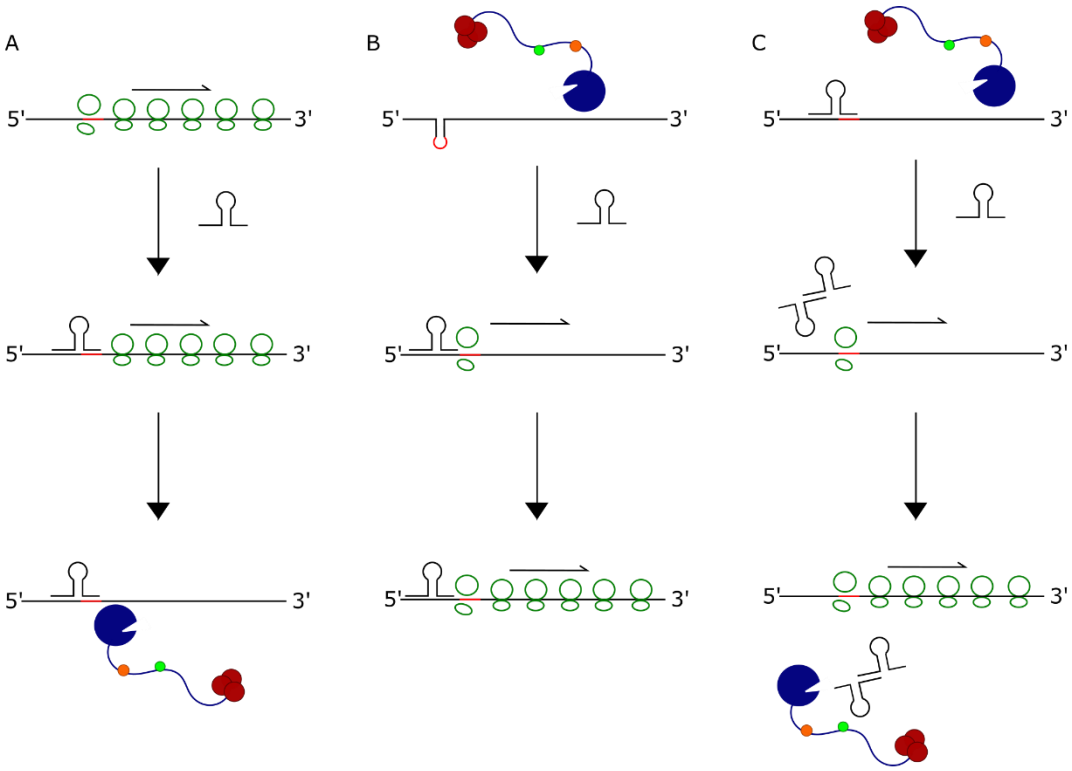


Figure 2: Schematic of the three most common sRNA functions. A) sRNAs most commonly act by binding to the Shine-Dalgarno box to repress translation initiation. mRNAs which are not protected by polysomes are vulnerable to degradation by the degradosome. Thus, by repressing translation, sRNAs also target mRNAs for degradation. B) mRNAs may have complex structures in the 5'UTR which prevents translation initiation. sRNA binding to the 5'UTR can remodel structured RNA and free the nucleotides necessary for translation initation so as to protect the mRNA from further decay as well as initiating the translation process. C) sRNAs can bind to other sRNAs to sequester them from their own function.

Global RNA regulation within prokaryotes has been shown to be vastly complex. Different sRNAs have been shown to participate in divergent pathways which diversely affect

mRNA half-lives. Some sRNAs such as RyhB bind to the ribosome binding site on mRNAs to block translation initiation (Figure 2A) (Chareyre & Mandin, 2018). Once translation initiation is blocked, polysomes can no longer form on the mRNAs which leaves them vulnerable to decay by RNase E and the RNA degradosome. Other sRNAs, such as ArcZ, DsrA, and RprA, can instead activate translation on their target mRNA (*rpoS*) upon binding (Repoila et al, 2003). The Shine-Dalgarno (SD) sequence of *rpoS* is buried in stem loop structures, preventing it from being accessed by ribosomes. ArcZ, DsrA, and RprA all have extensive sequence complementarity with the 5' region of *rpoS*. Upon sRNA binding, the SD sequence is released from the stem loop and may be accessed by ribosomes for translation initiation (Figure 2B) (Repoila et al, 2003).

In addition to the diverse methods by which post-transcriptional regulation is achieved with sRNAs, each sRNA may also have a variety of targets. GcvB, for example, has two distinct seed regions and is predicted to be involved in the regulation of up to 2% of all *E. coli* genes (Lalaouna, 2018). RyhB, another potent sRNA, can regulate many, if not all, of the proteins which may affect intracellular iron concentration, including proteins that regulate iron intake and non-essential proteins that sequester Fe^{2+} for functional reasons (Massé et al, 2007). The assortment of different modes of mRNA regulation by sRNA is only made more interesting by with the knowledge that most of these effects, however different, are mediated by Hfq and the RNA degradosome (De Lay et al, 2013). Both Hfq and the RNA degradosome directly bind RNA, but how they cooperate to achieve the multitude of effects of sRNA binding to mRNA is still an enigma, as well as how both Hfq and the RNA degradosome must be able to recognize the multitudes of structures adopted by the sRNA-mRNA complexes.

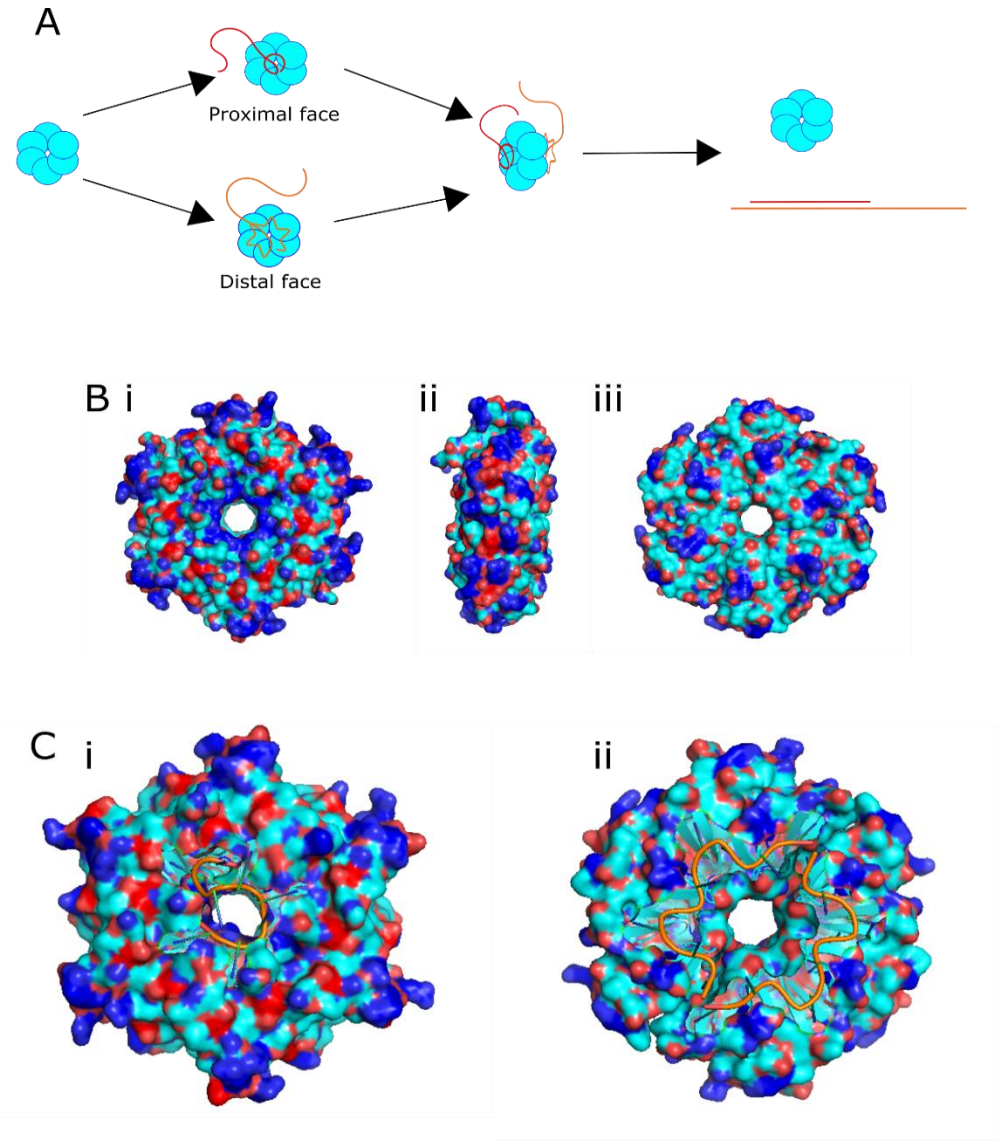


Figure 3: Hfq and its RNA binding modes. A) A schematic showing the annealing of an sRNA to its target mRNA with the help of Hfq. Generally speaking, sRNAs preferrentially bind the proximal face while mRNAs preferrentially bind the distal face. Hfq brings two RNAs in proximity to each other to aid in annealing. B) Surface structure of Hfq (PDB ID: 1HK9), with positively charged amino acids coloured in red and negatively charged amino acids coloured in blue. Subfigure i) shows the proximal face, ii) shows the rim, and iii) shows the distal face. C) The charge labelled surfaces bound to RNA. i) The proximal face binds to the RNA backbone through the positively charged rich in the center (PDB ID: 4VQS). Each Hfq monomer contacts one nucleotide. ii) The distal face binds to the RNA through the postively charged patches, which bind three nucleotides per monomer (PDB ID: 3GIB).

Hfq and RNase E, the scaffolding upon which the RNA degradosome binds, both contain general RNA-binding sites (Carpousis, 2007; Farner & Feig, 2013). Hfq is a homohexamer which contains multiple promiscuous RNA-binding sites, the most well studied of which are the proximal and distal faces. The details of these binding sites will be discussed in chapter 1 section 3 entitled Hfq Structure and Function. Briefly, the proximal face preferentially binds poly-U RNA sequences, which most frequently occur in the 5'UTR of mRNAs. The distal face, in contrast, preferentially binds sequences with the pattern of AAYAAYAA, where Y is any pyrimidine. The AAYAAYAA pattern more frequently occurs in sRNAs (Farner & Feig, 2013). Hence, it is easy to envision that Hfq can facilitate the annealing of sRNA to their target mRNAs by binding to both at the same time and eliminating the distance between the base-pairing regions (Figure 3A) (Vogel & Luisi, 2011). RNase E, the scaffolding protein upon which the degradosome assembles, contains two halves referred to as the N- and C-terminal domains (Figure 4C). The N-terminal domain is a structured endonuclease which prefers single stranded RNA with a 5'-monophosphate group. The C-terminal domain of RNase E contains microdomains used to bind other components of the degradosome or to provide additional support to the complex. These microdomains include, in order of N to C-terminal, the membrane targeting sequence (MTS), RNA-binding domain (AR1), RhlB-binding site (RBS), arginine-rich RNA-binding site (AR2), enolase-binding site (EBS), and PNPase-binding site (PBS). Details regarding RNase E and the RNA degradosome are discussed in the following section.

1.2 RNA Degradosome Evolution and Organization

At the heart of the various post-transcriptional regulators is the RNA degradosome, as both sRNAs and Hfq have both been shown to affect post-transcriptional regulation through influencing the activities of the degradosome. The RNA degradosome's importance for bacterial viability is doubly as evident when considering its homologues. In almost all species of bacteria, there exists some form of a degradosome (Aït-Bara & Carpousis, 2015). The ancient origin of RNase E has been suggested to predate the separation of proteobacteria, actinobacteria, and cyanobacteria (Lee and Cohen, 2003). Some degradosome like structures are also present in chloroplasts and mitochondria, which are thought to have evolved from cyanobacteria and alpha-proteobacteria respectively (Ochoa de Alda *et al*., 2014; Schein et al, 2008; Szczesny et al, 2012; Wang and Wu, 2015). Bacteria which do not have RNase E contain other RNases that fill the same function instead. For example, in B. *subtillis,* RNase E is replaced by RNase J and RNase Y (Lehnik-Habrink et al, 2011). Although not all these degradosomes are derived from the same evolutionary origin, the fact that similar complexes arose independently of each other points to the importance of such a complex for the survival and propagation of the bacterial species.

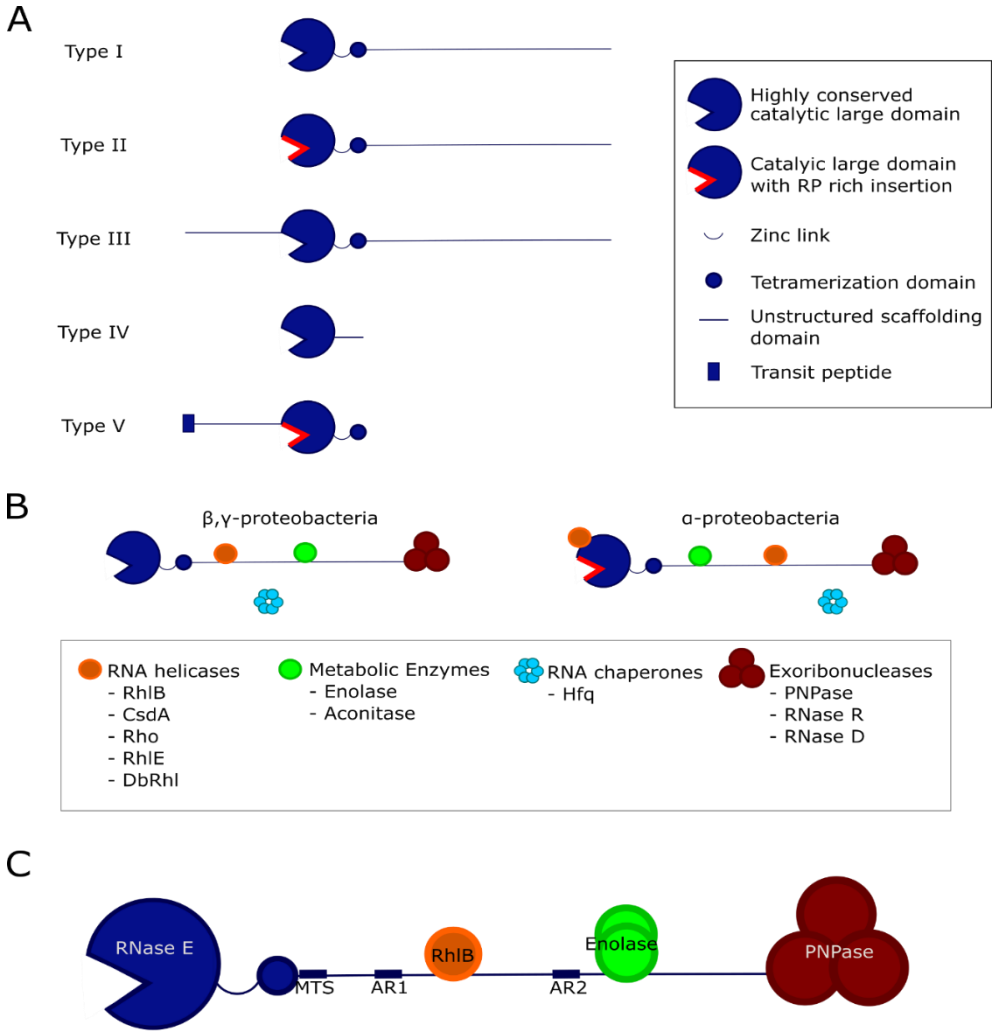


Figure 4: The organization of various kinds of degradosomes. A) The five types of degradosomes currently characterized. This schematic shows the cartoon structure of the scaffolding RNase, which can be RNase E, RNase G, RNase J, or RNase Y. All scaffolding RNases contain the highly conserved large catalytic domain. Most RNases contain a Zn link, the tetramerization domain, and an unstructured scaffolding region though the location of the scaffolding region may differ. B) The interaction partners of the degradosome and the proteins often associated with the degradosome in proteobacteria. Though Hfq does not have a binding domain on RNase E, it is often involved in the recognition and recruitment of the degradosome to targeted mRNAs. C) The organization of the E. coli RNA degradosome. The E. coli RNA degradosome consists of an RNase E, upon which the RNA helicase RhlB, the metabolic enzyme enolase, and the processive RNase PNPase binds. The MTS is the membrane targeting sequence which anchors the degradosome at the periplasmic membrane, and AR1 and AR2 are two RNA binding sites.

In bacteria, the RNA degradosome can take many forms, canonically categorized into five different types (Figure 4A) (Aït-Bara & Carpousis, 2015). All types of the degradosome contain an RNase upon which the body of the degradosome forms. These RNases share a similar catalytic core, which contains an RNase H-like domain, DNase I-like domain, S1 RNA-binding domain, as well as a Zn link which allows for dimerization (Figure 4A). Four out of five types also contain a small structured domain which allows for tetramerization. Alongside these structured elements, these RNases also possess an unstructured region containing various microdomains which recruit a variety of proteins that ultimately form the full degradosome complex (Figure 4B).

All degradosomes also contain a similar array of interaction partners which bind to the core RNase through its microdomains (Figure 4B). The presence of these microdomains are conserved throughout evolution, though their structure, sequences, and binding partners may all differ (Aït-Bara et al, 2015). For example, the PNPase-binding site on cyanobacteria show no sequence similarly to the PNPase-binding site in RNase E, even though they perform the same function (Aït-Bara et al, 2015). Although the compositions and organizations of the degradosomes are not identical, patterns can be gleaned from the evolutionary tree which may be used to ascertain the importance of each microdomain. The various degradosome's microdomains usually bind an array of similar accessory proteins such as helicases, metabolic enzymes, RNA-binding proteins, and exoribonucleases, though these accessory proteins may be different. *E. coli* RNase E, for example, binds the exoribonuclease PNPase while *P. syringae* RNase E binds the processive RNase R (Bandyra & Luisi, 2018; Purusharth et al, 2005). Similarly, the range of RNA helicases include RhlB, RhlE, Rho terminator, and more (Aguirre et al, 2017; Jager, 2001; Py et al, 1996). Metabolic enzymes which associate to the degradosome

include enolase, aconitase, acetyl transferase, and others (Chandran & Luisi, 2006; Hardwick et al, 2011; Kovacs et al, 2005). Though the specific role of these metabolic enzymes in the degradosome is unclear, their presence in the degradosome suggests a strong link between the abundance of cellular energy and RNA degradation. Interestingly, the composition of an individual degradosome can be flexible. By changing the growth or conditions surrounding the bacteria, different proteins can associate and dissociate from the degradosome. In *E. coli*, under cold shock conditions RhlB is replaced with the cold shock helicase CsdA, which helps facilitate more efficient RNA unwinding under less temperate conditions (Prud'homme-Généreux et al, 2004). Similarly, in *C. crescentus*, RhlB is replaced with RhlE under cold shock conditions (Aguirre et al, 2017), indicating that the formation of a "cold shock degradosome" is not limited to one species, and can potentially occur in all degradosome containing organisms.

1.2.1 RNase E

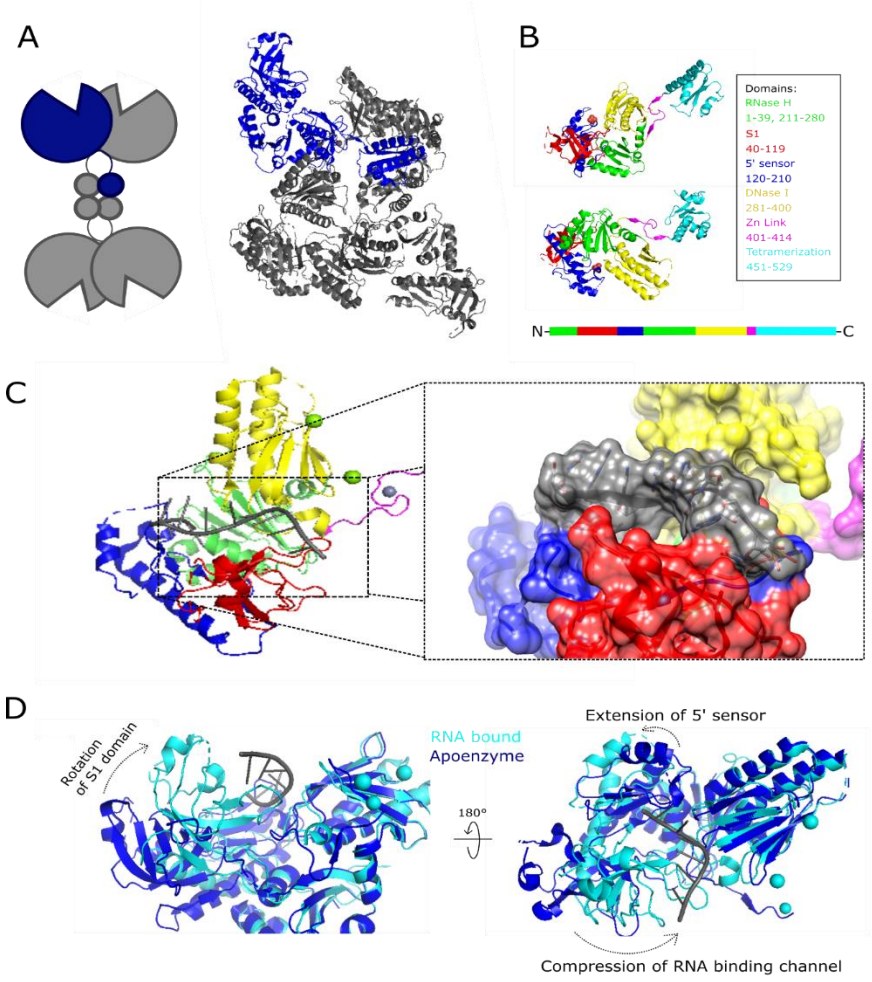


Figure 5: Structure of the N-terminal catalytic domain of RNase E. A) Tetramerization of RNase E (PDB ID: 2VMK). Each of the small tetramerization domain binds one other monomer and forms a dimer of dimers, with each dimer bound to one zinc ion through the Zn link. In addition to sharing a Zn ion, the RNase E protomers interact with each other extensively along the DNase I domains as well as the S1 domains. B) The 5 subdomains of the N-terminal structured domain. (PDB ID: 2VMK) C) The RNA binding channel. (PDB ID: 5F6C) The single-stranded RNA binds in a channel between the S1 and DNase I domains and contacts the 5' sensor at the end. D) The conformational changes induced upon RNA binding. (PDB IDs: 2VMK and 5F6C) The 5' sensor extends, pushing the S1 domain towards the DNase I domain which compacts the RNA binding channel and induces cleavage.

The most well characterized of all degradosomes is the E. *coli* RNA degradosome. The RNase E scaffolded degradosome is the canonical type I degradosome, which contains a core RNase — in this case, RNase E — with a structured catalytic N-terminal domain and an unstructured C-terminal tail upon which all other elements of the degradosome bind (Bandyra & Luisi, 2018). RNase E is organized as a tetramer, more specifically a dimer of dimers (Callaghan et al, 2005) (Figure 5A). Each monomer in a dimer cooperates structurally to cleave two pieces of RNA. The catalytic N-terminal domain of RNase E contains the RNase H-like domain, which contacts and stabilizes the RNA for cleavage; the DNase I like domain, which coordinates a catalytic Mg^{2+}; the S1 RNA binding domain, which forms a channel with the DNase I like domain and positions the backbone for cleavage; the structural Zn link and the small tetramerization domain, both of which work together to coordinate tetramerization. The N-terminal domain contains a small channel formed by the S1 and DNase I like domains which only fits a single stranded RNA (Figure 5C). Though RNase E has shown no sequence specificity, it prefers AU rich regions and prefers to cut two nucleotides downstream of a guanosine base (McDowall et al, 1994; Redko et al, 2003). Additionally, RNase E's endonucleotylic activity is enhanced upon binding of a 5'-monophosphate (Jiang & Belasco, 2004). Binding of a 5'-monophosphate is thought to induce a conformational change in the S1 domain of one dimer which compresses the catalytic channel and optimally positions the RNA backbone towards the catalytic Mg^{2+} of the other dimer (Koslover et al, 2004) (Figure 5D). As RNase E is also required in the cleavage of complex RNAs which do not contain a 5'-monophosphate, current understanding postulates two pathways by which RNase E cleaves RNA — the 5'-end dependent pathway and the 5'-end independent pathway, also known as direct entry pathway (Bouvier & Carpousis, 2011). The 5'-end dependent pathway requires the

presence of a 5'-end monophosphate, either from the mRNA substrate itself or from a different strand of sRNA. Indeed, studies have shown that, in the absence of a 5'-monophosphate on the target mRNA, sRNAs with a 5'-monophosphate can bind to the 5'-end sensor in RNase E and activate the cleavage activity in trans (Bandyra et al, 2012). The direct entry pathway, however, is less well understood and is thought to involve activation of the RNase activity of RNase E with the tertiary structure of the RNA itself (Baker & Mackie, 2003; Joyce & Dreyfus, 1998). Genome wide surveys have shown that the direct entry pathway may well be the major cleavage pathway in RNase E (Clarke et al, 2014).

The C-terminal of RNase E is equally complex, as it contains many microdomains which are essential for the formation of the degradosome (Bandyra & Luisi 2018). The majority of the C-terminal domain is unstructured except for the conserved structured regions which serve specific functions (Callaghan et al, 2004). These structured regions include the membrane targeting sequence (MTS), arginine-rich regions 1 and 2 (AR1 and AR2) which serve to bind RNA, the RhlB-binding site (RBS), enolase-binding site (EBS), and PNPase-binding site (PBS) (Figure 4C) (Callaghan et al, 2004; Chandran & Luisi, 2006; Khemici et al, 2008). The MTS is an amphipathic helix that tethers the degradosome onto the membrane (Khemici et al, 2008). Upon membrane binding, RNase E seems to be stabilized and affinity for the substrates seems increased (Murashko et al, 2012). It has been suggested that membrane localization of the degradosome machinery, in addition to stabilizing the complex, creates a temporal delay between the transcription of new mRNAs and their degradation to allow for mRNAs to be translated (Mackie, 2013). Fluorescent microscopy has shown that RNase E is equally distributed throughout the membrane but forms local foci which can be disrupted by the addition of rifampicin (Strahl et at, 2015), which suggest that perhaps substrate localization

affects degradosome localization as well. Though the quaternary structure of the degradosome is known, due to the difference in stoichiometry of RNase E and PNPase, large scale organization of the degradosome is still unknown but suspected to be complex and mobile due to the flexibility and mobility of the C-terminal domain.

1.2.2 RNA helicase B

One of the components of the *E. coli* degradosome is RNA helicase B (RhlB) (Chandran et al, 2007). The majority of the cellular RhlB is thought to be bound to the degradosome and aids in unwinding complex RNA structures (Domínguez-Malfavón et al, 2013). RhlB is part of the DEAD box family of RNA helicases which have diverse functions in most organisms (Cordin et al, 2006). Notable DEAD box helicases include the eIF4A translation initiation factor, Prp28 and Prp5, which are involved in mRNA splicing (Xing et al, 2019). Evolutionarily speaking, modern helicases are derived from an ancestor with a conserved catalytic core similar to the helicase cores found today. Structurally, the DEAD box helicase cores contain two RecA like domains, which in turn contains eight conserved motifs, making up the ATP-binding domain and helicase C-terminal domain (Figure 6C) (Linder et al, 1989). The eight conserved motifs are the Q-motif, motif 1, 1a, 1b, motif II, III, IV, and V. These motifs are highly conserved and implicated in both ATP-binding and RNA association. The ATP-binding motifs consist of the Q motif, motif 1, 1a, 1b, and motif II (Tanner et al, 2003), and the RNA binding motifs consists of motif III, IV, and V. Motifs 1 and II are also commonly known as the Walker A and B motifs respectively. The Walker B motif contain the residues Asp-Glu-Ala-Asp (DEAD) which is essential for ATP hydrolysis and gives this family of proteins its name.

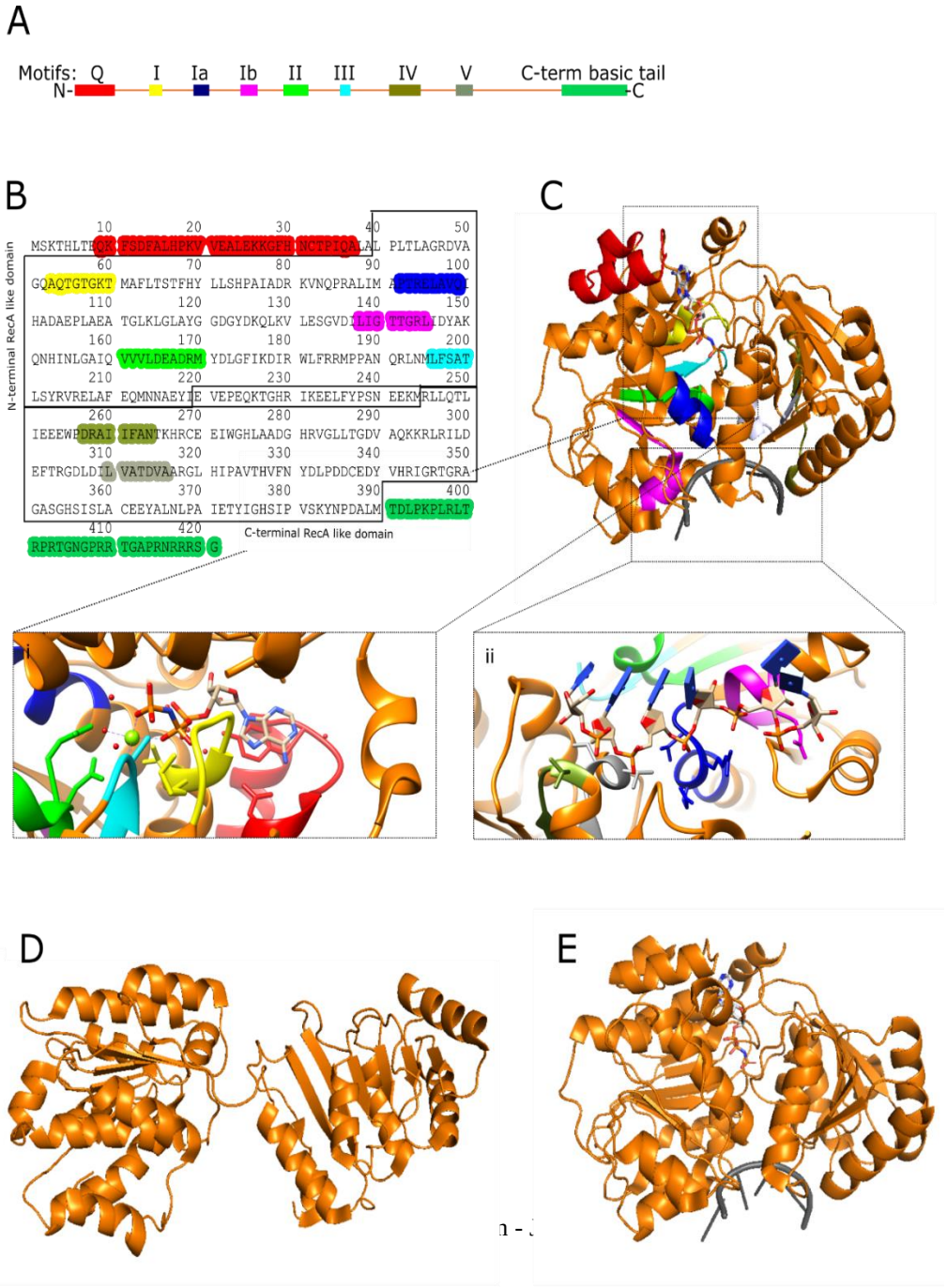


Figure 6: RhlB and DEAD box helicase structural organization. A) The eight motifs in RhlB plus the addition of the basic tail which is unique to RhlB. B) The sequence of RhlB coloured by motif. C) The canonical DEAD box protein core of an RhlB homologue (Vasa) with the corresponding motifs coloured (PDB ID: 2DB3). A closeup of the ATP and RNA binding motifs are shown in subfigures Ci and Cii. Motifs Q, 1, II, and III are shown to contact the ATP directly while motifs 1a, 1b, IV, and V contact RNA. D) and E) shows the movement between the two RecA like domains and how, upon binding to ATP and RNA, the two domains compact. (PDB IDs: 5GI4 and 2DB3)

Though every DEAD-box helicase shares the same core, each individual helicase is also adapted to their specific cellular functions, which vary significantly in the cell (Linder et al, 1989). Most commonly, DEAD-box helicases are responsible for the disassembly of ribonucleoproteins, chaperones for RNA folding, and stabilizing of protein-RNA interactions (Jarmoskaite & Russell, 2011; Linder & Jankowsky, 2011; Weston & Sommerville, 2006). For example, Mss116 and CYT19 act as non-sequence specific RNA chaperones, Sub2 promotes protein-protein interactions on RNA, and Dbp4 is involved in the release of snoRNA (Xing et al, 2019). When DEAD-box helicases contain modifications or auxiliary domains, they are always specific to their individual functions and come from N and C-terminal additions without disrupting the central core. In such a sense, RhlB also contains additions to the helicase core that are specialized for its function. RhlB contains an unstructured C-terminal tail enriched with positively charged residues, theorized to aid in RNA binding (Figure 6B) (Chandran et al, 2007). In addition, upon binding to RNase E, the ATPase activity of RhlB is stimulated (Chandran et al, 2007). As ATPase activity is required for the release of the substrate, subsequently freeing the helicase for more activity, stimulated ATPase activity leads to faster unwinding (Chandran et al, 2007). These characteristics together support the idea that RhlB is optimized for aiding RNA unwinding and remodelling in the *E. coli* RNA degradosome (Carpousis 2007). RhlB's role in the degradosome is seemingly straightforward yet its associations with different members of the degradosome are ill-described. Though previously known to be implicated in the processive degradation process by PNPase, recently, the RhlB binding site on the degradosome has been shown to be required for Hfq binding in certain circumstances (Ikeda et al, 2011; Khemici & Carpousis, 2004). As Hfq is not thought to be

involved in the processive degradation pathway of the degradosome, RhlB's role purely as an unwinding accessory to PNPase has been put into question.

1.2.3 Enolase

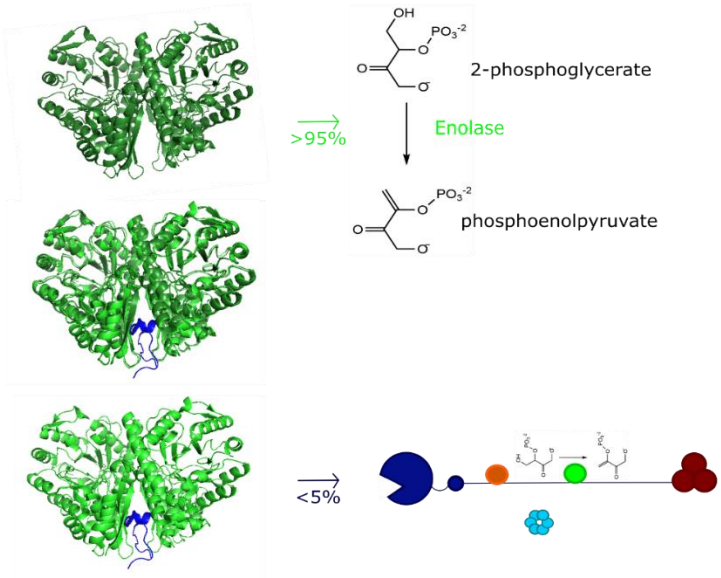


Figure 7: Enolase and its associations with the *E. coli* RNA degradosome. Most of the enolase in the cell is not attached to the degradosome but rather actively participating in glycolysis. However, a small percentage of the enolase present is attached. Attachment to the degradosome confers very little change in conformation. On the left hand side three structures of enolase are shown. From top to bottom they are free enolase (PDB ID: 1E9I), free enolase overlaid with enolase bound to RNase E (PDB IDs: 1E9I and 2FYM), and enolase bound to RNase E (PDB ID: 2FYM) respectively.

Another protein, a portion of which is associated with the degradosome is enolase, a glycolytic enzyme. Enolase's primary role in the cell is to convert 2-phosphoglycerate into

phosphoenolpyruvate in the penultimate step of glycolysis (Figure 7). As such, enolase is abundant in the cytoplasm and only <10% of all enolases are associated to the RNA degradosome (Nurmohamed et al, 2010). The exact role of enolase on the RNA degradosome has largely remained a mystery, but some clues to its role have been collected over the years. In many degradosome homologues, an enzyme in the energy pathway has been found to associate with the degradosome. Some of them are in the glycolytic pathway (such as enolase), and some of them are found in the citric acid cycle (such as aconitase) (Hardwick et al, 2011). Thus, the association of an enzyme in the energetic pathway to the degradosome seems to be beneficial to the survival of the species. Additionally, sRNA-mediated degradation of *ptsG* has been shown to have decreased functionality upon removal of the enolase binding site in the degradosome (Morita et al, 2004). *ptsG* encodes a major glucose transporter responsible for importing glucose. Thus, *ptsG* regulation is implicated in energy conservation and is therefore intimately related to the glycolysis cycle. As the enolase binding site in the scaffolding region is adjacent to the RNA binding site AR2, binding of enolase could hypothetically alter the structure of the AR2 binding site and change its preference and affect *ptsG* degradation. However, little structural information is available to validate such a hypothesis. To further understand the function of enolase in the context of the degradosome, more structural and functional data is needed.

1.2.4 Polynucleotide Phosphorylase

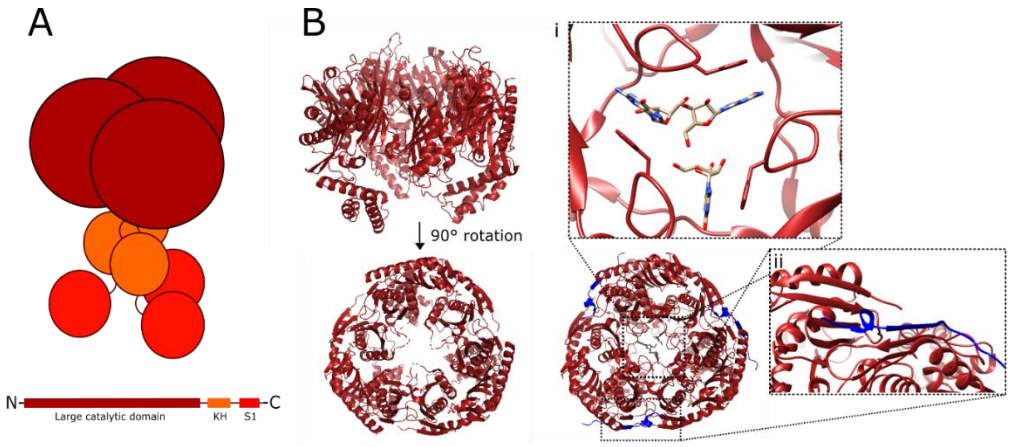


Figure 8: PNPase organization, RNase E and RNA binding. A) Schematic of the quarternary structure of PNPase. PNPase is formed of three monomers each containing three domains, the large catalytic domain, KH and S1 domains. The large catalytic domains form a ring (B) in which the RNA in grey binds (i) (PDB ID: 3CDI). PNPase associates with RNase E in dark blue through the interace of its trimers (ii) which make contact with two monomers at the same time. (PDB ID: 3GCM)

One of the most important proteins which associates with the degradosome is polynucleotide phosphorylase (PNPase), which is the processive RNA degradation center of the degradosome (Carpousis, 2007). PNPase functions in the degradosome as a 3'-5' exoribonuclease by using an inorganic phosphate to attack the phosphodiester backbone linkage at the 3' end, generating a diphosphate nucleotide (Nurmohamed et al, 2009). Similarly to enolase, PNPase itself also exists independently of the RNA degradosome (Del Favero et al, 2008). Only a small fraction of cellular PNPase is bound to the degradosome at any time. In the cell, PNPase forms a trimer which is its functional oligomeric state (Figure 8A). Each PNPase monomer also contains three domains, the RNase PH-like domain, KH domain, and S1 domain (Nurmohamed et al, 2009). The RNase PH-like domain is the catalytic domain and the most well characterized of the three. A trimer of this domain is analogous to the eukaryotic exosome,

where the three PNPase PH domain forms a ring, the inside of which is a channel through which RNA passes. The RNA that passes through the channel is then degraded into diphosphate nucleotides. The more enigmatic KH and S1 domains have been theorized to aid in RNA recruitment, and more recently even shown to have possibly an RNA chaperoning role (Cameron & De Lay, 2016). In its association with RNase E, PNPase binds to the degradosome through beta sheet interactions with the conserved microdomain on the RNase E C-terminus (Nurmohamed et al, 2009). As PNPase exists as a homotrimer and RNase E exists as a homotetramer, the stoichiometry of PNPase and RNase E may create a complex web of interactions along the membrane, the details of which are still obscure.

In this thesis, the main protein of focus will be RhlB. To investigate RhlB function within the degradosome, I elected to use a subassembly of the RNA degradosome henceforth referred to as the ternary complex. The ternary complex consists of RNase E 603-850, RhlB, and enolase (Figure 9). RNase E 603-850 comprises of two RNA-binding sites (AR1 and AR2) along with the RhlB and enolase-binding sites. As RNase E binding stimulates RhlB ATPase activity (Chandran et al, 2007), RhlB in complex with RNase E will provide a more accurate measurement of RhlB activity in the context of the degadosome in comparison to free RhlB. In addition to including the RNase E segment, enolase has also been included for its potential to modulate the functional capacity of certain sRNAs (Morita et al, 2004). To simplify the purification process and eliminate excess complexity, the RNase E N-terminus including the MTS was not included. For similar reasons, PNPase was also not included in the studies. The three chosen components together will hopefully represent a significant portion of the degradosome C-terminal tail which play a large role in RNA binding and substrate selection.

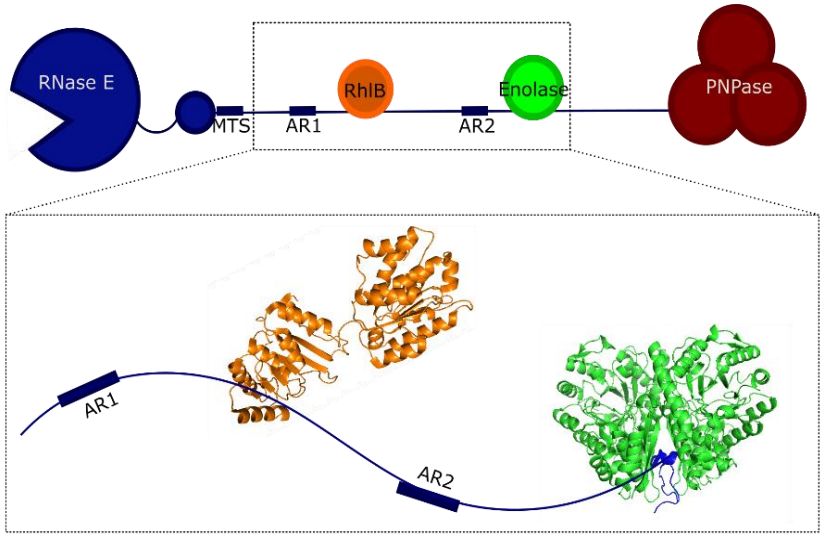


Figure 9: Cartoon representation of the ternary complex. Three elements of the degradosome were selected as the focal point of this thesis. These proteins were chosen to represent a significant portion of the C-terminal domain. The three selected elements are RNase E 603-850 shown in blue, RhlB shown in orange (modelled using MODELLER (Webb & Sali, 2016)) , and enolase shown in green. (PDB ID: 2FYM)

1.3 Hfq Structure and Function

One protein which is deeply involved in global RNA regulation is Hfq. Hfq was originally discovered in 1968 to be essential for the replication of bacteriophage Qβ (de Fernandes et al, 1972). It was thus named Hfq for Host Factor for bacteriophage Q beta. Since then, Hfq has been found in many species of bacteria, and its homologues have been found in eukaryotes and archaea as well (Hermann et al, 1995; Törö et al, 2001). Hfq belongs to a family of promiscuous RNA-binding proteins called the Sm/LSm proteins, which are an ancient family of proteins theorized to have evolved from a single ancestor from before the split of the kingdoms eukarya and archaea due to its ubiquitous presence (Hermann et al, 1995). These

proteins were first discovered as an antigen to the anti-Sm antibodies found in patients with systemic lupus erythematosus (Tan & Kunkel, 1966), but various other proteins similar to the structure of these antigens were later found to exist across all domains of life. These proteins all share a characteristic fold and oligomerization which defines the family but they all participate in cellular function differently. Members of the Sm/Lsm proteins include those found in the spliceosome as ribonucleoproteins, as RNA chaperones for snRNAs, or as RNA remodelling proteins (Schümperli & Pillai, 2004; Woodson et al, 2018). These proteins exhibit a characteristic fold and ring-like quaternary structure. Two motifs are found in these proteins, Sm1 and Sm2 (Hermann et al, 1995). Sm1 contains the N-terminal alpha helix and beta strands one to three, while Sm2, contain beta strands four and five. Beta strands 3-5 fold in on themselves, with beta strands 4 and 5 forming the protomer interfaces during oligomerization (Figure 10). Between eukaryotes and prokaryotes, the largest differences between these proteins come from the Sm2 motif (Kambach et al, 1999). The C-termini of the Sm/Lsm family of proteins can also differ in length, but these elements are less well described. The monomers usually oligomerize into a ring of six to seven units with two distinct RNA binding faces (Figure 3B, Figure 10) (Bass et al, 2002). These oligomers could be either homomeric or heteromeric, with the heteromeric rings consisting of two different kinds of monomers spaced out evenly (Zhou et al, 2014). The specific binding capacities of each protein depends on the Sm2 motifs of the monomers which form these rings (Hermann et al, 1995).

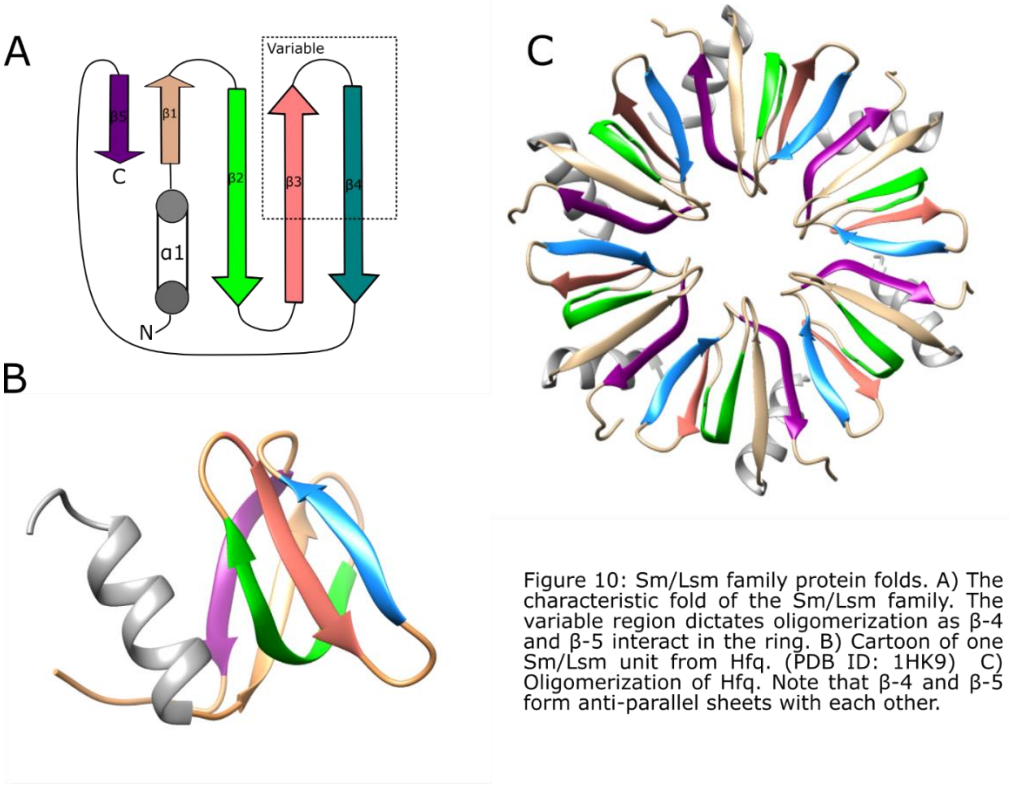


Figure 10: Sm/Lsm family protein folds. A) The characteristic fold of the Sm/Lsm family. The variable region dictates oligomerization as β-4 and β-5 interact in the ring. B) Cartoon of one Sm/Lsm unit from Hfq. (PDB ID: 1HK9) C) Oligomerization of Hfq. Note that β-4 and β-5 form anti-parallel sheets with each other.

In *E. coli*, Hfq forms a homohexameric ring which contains multiple RNA-binding sites, the proximal site, distal site, and rim site (Figure 3B) (Hoekzema et al, 2019; Panja et al, 2013; Sauer & Weichenrieder, 2011). The proximal face of Hfq is the surface in which the N-terminal alpha helices of all the monomers are situated, whereas the distal face is made up of various residues from the beta-strands (Sauter et al, 2003). The C-terminal tail of Hfq contains many acidic residues and extends out from the rim of the ring in an unstructured manner. The proximal face of Hfq makes contact with RNAs through positively charged residues close to the centre

of the ring's aperture (Figure 3C) (Wang et al, 2011). Each Hfq monomer would make contact with the backbone of one nucleotide. The base pairing faces of each nucleotide, when bound to Hfq, would be exposed to the solvent and accessible for binding. The proximal face of Hfq seems to prefer U rich sequences commonly found in sRNAs (Vogel & Luisi, 2011). In contrast, the distal face of Hfq interacts with RNA in a very different way. Each Hfq monomer makes contact with three RNA bases on the distal face (Figure 3C) (Link et al, 2009). Two bases are secured while one is exposed to the solvent and free for base pairing. The distal face prefers A rich RNA regions which are more commonly found in the UTRs of mRNAs (Vogel & Luisi, 2011). In this way, Hfq facilitates the annealing of sRNAs to their target mRNAs by bringing the two RNAs in proximity to each other, with the sRNA bound to the proximal face and the mRNA bound to the distal face. Though these modes of binding are the most commonly described, Hfq has also been shown to bind RNA on both sides simultaneously, to bind DNA and to bind RNA on the rim of the hexameric ring as well (Orans et al, 2017; Panja et al, 2013; Schu et al, 2015). These methods of binding have yet to be described in depth and their functions in vivo await further studies.

In vivo studies have shown that Hfq is limiting in the cell when compared to RNA, perhaps unsurprisingly (Santiago-Frangos & Woodson, 2018). The large number of RNA in the cell coupled with Hfq's promiscuity means that Hfq is constantly sequestered by one RNA or another. In addition, many sRNAs have been shown to require Hfq for their function and thus sequester the chaperone from other RNAs. sRNAs by themselves are very vulnerable to degradation and thus they require either binding to their target RNA or binding to a protein for stabilization. Hfq fulfils this role for sRNAs, where it protects the RNA from degradation and also shuttles it to its target mRNA (Vogel & Luisi 2011). There are several ways in which Hfq

can fulfill its role as the RNA chaperone aside from increasing the half-life of sRNAs. Firstly, it can increase rates of sRNA and mRNA annealing. In vitro rates of sRNA and mRNA annealing without a chaperone are too slow to have any biological relevance (Kawamoto et al, 2006; Soper & Woodson, 2008). However, upon Hfq binding, the seed region of the sRNA (the mRNA annealing segment) comes into proximity with the binding site on the mRNA, effectively reducing the annealing time (Zhang et al, 2002). In this way, Hfq acts in a similar way to a catalyst, as it lowers the entropic cost afforded to the "substrates" to facilitate a faster reaction. Another of Hfq's actions is to create energetically stable base-pairs. mRNA UTRs and sRNAs usually contain higher order structures which are intrinsically stable and difficult to penetrate. Thus, though transient sRNA-mRNA pairings are frequently formed without Hfq, they may have difficulty forming the most stable duplex as intermediate duplexes may dissociate too soon to be biologically relevant. Hfq binding ensures that the most energetically stable base-pairing is achieved and released, allowing the duplex to be fully functional in the cell. Lastly, Hfq can remodel the structures of the sRNA and mRNA, allowing for base pairing to occur (Soper et al, 2010). 5'-UTRs on the mRNA as well as sRNAs are usually structured, providing the RNAs with protection against ribonucleases. However, many of the sRNAs seed regions and its target regions on the mRNA are therefore buried under secondary structures and inaccessible for base pairing. Access to these elements may require more energy than the ambient energy available to sRNAs. Hfq then can provide an extra layer of interactions which can compete with intrinsic structures alongside sRNA-mRNA interactions (Geissmann & Touati, 2004; Hoekzema et al, 2019; Zhang et al, 2002). Hfq can position the seed regions and target sequence optimally to promote base pairings which otherwise may be difficult to form.

In this way, Hfq again acts similarly to a catalyst, lowering the required enthalpy for duplex formation.

One question that permeates Hfq-RNA interactions is competition. As Hfq is a limiting factor within the cell, it must be recycled from currently bound RNAs for efficient function (Santiago-Frangos & Woodson, 2018). Currently, there are two models that can explain Hfq-recycling in vivo. The first model is the degradation model, where in vivo, RNase E and the RNA degradosome are continuously degrading mRNA which is bound to sRNA and Hfq (Massé et al, 2003). Thus, once the mRNA and sRNA are degraded, Hfq is free to bind to new RNA partners. The second model is the competition model, where one binding RNA is continuously displaced by other Hfq binding RNAs (Fender et al, 2010). In this model, the RNA sequence does not hold the most importance, as Hfq is highly promiscuous. However, the concentration of different RNA species in the cell will make the greatest impact on the population of RNA bound to Hfq. One can envisage that these two models occur simultaneously, where Hfq is continuously liberated by RNase E and degradosome degradation while the RNA population in the cell is constantly titrating each other from Hfq, thus allowing the cell to rapidly adapt to changing environments.

More recently, deeper investigations into Hfq function have attempted to link structural elements to the functional patterns seen in various binding assays as described above. Most notably, Hfq's unstructured acidic tails have been shown to have an autoinhibitory function (Santiago-Frangos et al, 2019; Santiago-Frangos & Woodson, 2018). The acidic tails can mimic the phosphodiester backbone and sweep across the rim, displacing any weakly binding RNA by electrostatic repulsions. Like any unstructured regions, divergence of the Hfq tails between species is more pronounced than structured regions. However, in most species of Hfq, though

tail length varies significantly, similar acidic residues can be found, suggesting that the tails are indeed functionally relevant (Santiago-Frangos et al, 2019). In addition, more evidence of RNA remodelling by Hfq has been published. *dgcM* mRNA has been shown to be regulated by the sRNAs OmrA/OmrB (Hoekzema et al, 2019). Without Hfq, the 5' end of *dgcM* base pairs with a downstream segment which makes a large loop that contains two stem-loops itself (Figure 11). The first stem-loop sequesters the start codon while the downstream stem-loop shortens the sRNA binding region. Upon only Hfq binding, the stem-loops open, allowing for sRNA binding, thus committing the mRNA to the fate of degradation (Figure 11). Recent evidence has also been collected for Hfq binding to the bacterial chromosomal DNA (Orans et al, 2017). The proposed function of this mode of binding is to compact the DNA and organize it. However, more in vivo evidence is needed to ascertain whether this phenomenon is simply a physical association or if it indeed has physiological implications.

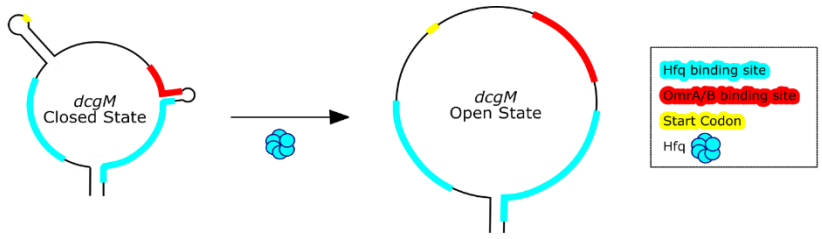


Figure 11: Schematic of *dcgM* 5'UTR remodelling by Hfq. Without Hfq, *dcgM* 5'UTR sequesters the binding site of OmrA/B, rendering them ineffective. When Hfq binds to the stem-loop structure, the 5'UTR region opens and allows for sRNA binding. Upon sRNA binding, *dcgM*'s translation is stalled and the mRNA is targeted for degradation.

Though well described through functional studies, structural studies of Hfq with sRNAs or mRNAs are few. To date, there are only three crystals structures of Hfq in complex with RNA, two of which used synthetic RNAs designed to bind a specific face of Hfq (Dimastrogiovanni et al, 2014; Link et al, 2009; Wang et al, 2011). Physiologically, the

interactions between sRNAs and Hfq are more diverse than what current structural data shows. sRNAs bind Hfq in a variety of different ways and can be classified based on their Hfq binding mode. Class I sRNAs can bind Hfq through the proximal face as well as the rim, while class II sRNAs bind Hfq through the proximal and distal faces (Updegrove et al, 2017). Only one structure of an sRNA bound to Hfq has been experimentally determined. Dimastrogiovanni et al published the crystallographic structure of RydC bound to the proximal face of Hfq in 2014. The structure of both the distal face and rim binding sites of Hfq bound to native RNA is therefore currently undescribed. The tails of Hfq are also highly unstructured and mobile, and thus their mechanisms of action, though biochemically understood, is structurally obscure. In addition, mRNA binding modes of Hfq have also eluded structural determination, as these RNAs are much larger than sRNA. Thus, though we understand the mechanism behind Hfq's function, its finer details remain mysterious.

One of the most important protein partners of Hfq is the RNA degradosome. Hfq has been shown to recruit the RNA degradosome when bound to an mRNA and its target sRNA (Bandyra et al, 2013). However, the method by which Hfq interacts with the degradosome is still unclear. Without the RNA degradosome, sRNAs and Hfq cannot carry out their respective functions as the RNA degradosome is the effector of many sRNA functions by degrading the sRNA target of repression. Unsurprisingly perhaps, most of the cellular Hfq is in proximity to the membrane (Diestra et al, 2009). Localization of Hfq to the membrane could potentially support Hfq's close functional association with the RNA degradosome, which is itself tethered to the membrane. Hfq has been shown to bind to the C-terminal of RNase E (Ikeda et al, 2011). Residues 711-750 of RNase E are required for *ptsG* mRNA degradation, which also requires Hfq mediation. The peptide from residues 711-750 can bind Hfq independently of RNA

presence, suggesting that Hfq can associate transiently with RNase E intracellularly by itself (Ikeda et al, 2011). More interestingly, recent studies have suggested that Hfq can also bind to the less well studied KH and S1 domains of PNPase (Bandyra et al, 2016). Contrary to the degradative function of PNPase, when bound to the KH and S1 domains, PNPase and Hfq confer more protection to sRNAs than sRNAs bound to Hfq alone. Without Hfq, however, PNPase will still degrade the very same sRNAs it once protected. These interactions highlight the dynamic way in which Hfq interacts with the RNA degradosome for function and also the vast amount of information still unknown about Hfq's dynamic role in global RNA regulation.

1.4 Small Regulatory RNAs: RyhB, SgrS, and ChiX

In *E. coli*, sRNAs comprise a class of RNAs from around 40 nt to 500 nt in length (Viegas & Arraiano, 2008; Vogel & Wagner, 2007). According to Regulon DB, there are currently 117 sRNAs and 235 sRNA mediated genomic interactions which are described to date (Santos-Zavaleta et al, 2019). Most of these interactions have only been discovered within the last decade. As better sRNA identifying software emerge, many sRNAs and their interaction partners will no longer remain hidden and thus require further research for a thorough understanding. sRNAs not only have a variety of lengths and structures, but they also have a variety of mechanisms of action (Figure 2). Most sRNAs base pair to mRNAs and target it for degradation, but sRNAs can also protect mRNAs and aid in translation initiation as well (Kavita et al, 2018). Most commonly, sRNAs will base pair with the mRNA next to the Shine-Dalgarno sequence in the 5'-UTR to obscure the ribosomal binding site and prevent translation initiation (Massé et al, 2003). As translation initiation cannot occur, polysomes thus cannot form on these

naked mRNAs, making them vulnerable to attack by RNases. Similarly, sRNAs can base pair to mRNAs downstream of the translation initiation site, stalling initiation and causing degradation (Pfeiffer et al, 2009). Interestingly, sRNAs can also change the population of translated proteins within an operon (Desnoyers et al, 2009). Some sRNAs have been found to bind in the UTRs between two proteins-coding regions in one operon, causing translation to be halted for one protein but not others (Papenfort et al, 2013). Other times, mRNAs may protect themselves from degradation by hiding their RNase E cleavage site in structured loops (Desnoyers et al, 2009). sRNA binding to these structured loops can expose RNase cleavage sites which are otherwise inaccessible in the native conformation, rendering the mRNAs vulnerable or can create “alternative splicing” for prokaryotes (Desnoyers et al, 2009). At the other end of the spectrum, sRNAs can also activate translation. Some mRNAs bury their Shine-Dalgarno sequence in secondary structure, thereby preventing ribosomal access. sRNAs, with the aid of Hfq, can remodel complex RNA structures, thus freeing up sequences required for ribosome binding and translation initiation (Morfeldt et al, 1995; Prévost et al, 2007). Not only do sRNAs regulate mRNA lifetimes and translations, sRNAs may also regulate the functions of other sRNAs as well (Lalaouna et al, 2015). LeuZ functions as a sRNA sponge to base pair with at least two sRNAs and target them for degradation once bound, adding another layer to post-transcriptional regulation by sRNAs. Similarly to LeuZ, sRNAs can posesss more than one target RNA, and can exert different effects on them as well (Lalaouna et al, 2018; Wang et al, 2015). Hence, the web of RNA-mediated post-transcriptional regulation is vastly complex. Deeper research needs to be done to further understand how the cellular population of sRNAs are made and respond to various internal and external signals.

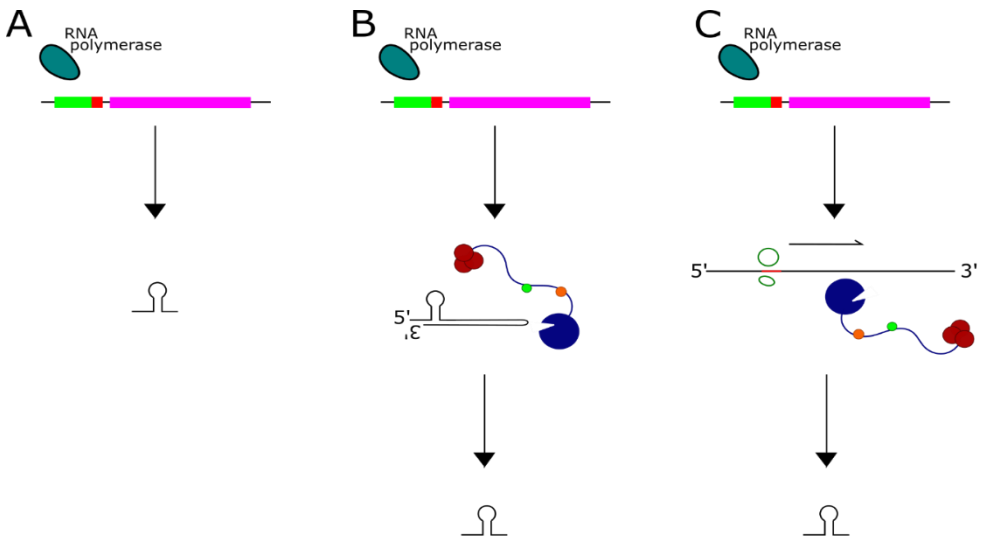


Figure 12: Schematic of the most common ways of sRNA maturation. A) A sRNA can be directly transcribed from its own operon in its active form. This method of sRNA creation does not require any secondary processing. B) sRNAs can be transcribed in an inactive form, then cleaved by RNase E to make an active form. C) sRNAs can be cleaved from the UTRs of mRNAs.

sRNAs are not only diverse in function, they are also diverse in the ways they are created (Figure 12). Many sRNAs contain their own ORFs which allows for independent transcription regulation (Antal et al, 2005; Delihas & Forst, 2001; Suzuki et al, 2015). For other sRNAs, the transcribed RNA is longer than the functional sRNA and needs to be cleaved for full functionality (Moores et al, 2017). For example, DrrS is a sRNA in *Myobacterium tuberculosis* which functions as a 109 nt RNA. However, researchers noted that the 3'-terminal of the RNA cannot efficiently terminate by itself or recruit termination factors. At the same time, repetition of the 6nt at the 3'end of the physiologically active RNA led to elongation of the functional RNA, which suggested that these nucleotides formed a cleavage site. These observations led to

the discovery of the full transcript, DrrS+, which is heterogenous in length with the most prominent species at around 273 nt long. Drrs+ thus must undergo cleavage to become its mature form of at 109 nt in length.

Aside from sRNAs with their own transcriptional regulators, sRNAs can also be processed from the UTRs of mRNAs (Figure 12C) (Chao & Vogel, 2016). One recent example is found in the Cpx pathway, which is the pathway implicated in the stress response to cell envelope protein misfolding. CpxQ, a 60 nt sRNA, is released from the 3'-UTR of the mRNA *cpxP* during degradation (Chao & Vogel, 2016). The protein CpxP acts as a chaperone to misfolded outer membrane proteins, while the sRNA CpxQ negatively regulates expression of outer membrane proteins such as NhaB (sodium porin), fimA (fimbrin filament implicated in bacterial attachment to host cells), and Skp (outer membrane chaperone protein) (Chao & Vogel, 2016). The inclusion of sRNAs in UTRs give cells more flexibility and control of protein translation. In this case, two regulatory factors are derived from the mRNA *cpxP*, one protein and one RNA, both exerting the same pressure on the cell but at two locations, the outer membrane and cytoplasm. This property makes one mRNA doubly as effective while using less resources, making the cell more responsive to stimuli.

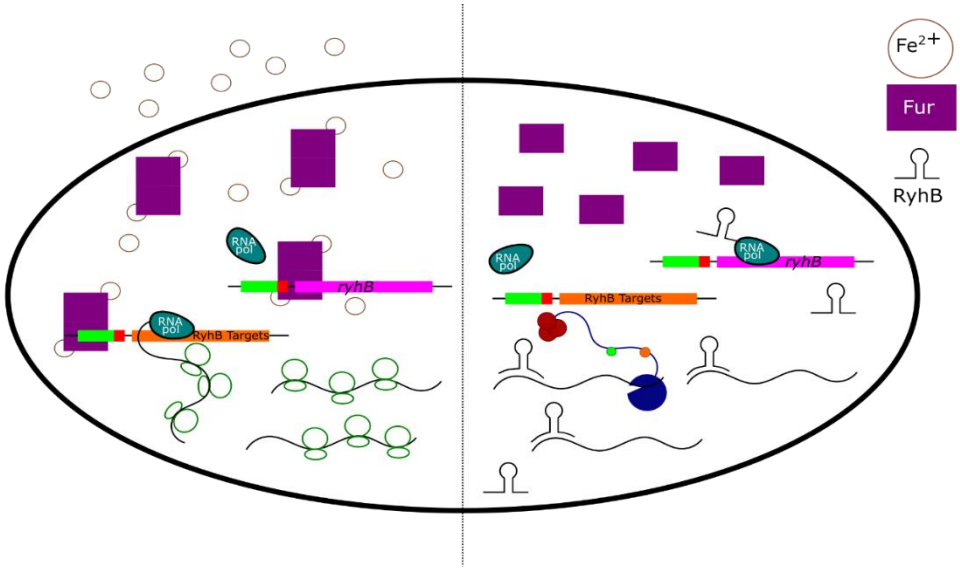


Figure 13: Simplified schematic of iron stress response in the cell. On the left, biologically active iron is abundant and binds to Fur (Ferric Uptake Regulator), the iron sensing protein. Fur is a transcriptional factor which represses RyhB transcription and siderophore synthesizing proteins when bound to free Fe^{2+}. In contrast, when iron concentrations are low, Fur is inactive. RyhB is thus transcribed and acts on mRNA such as *sodB* (superoxide dismutase) to repress translation and degrade the mRNAs. RyhB's targets contain iron binding proteins which are non-essential to survival. Thus, halting translation of these proteins can allow the cells to redistribute what little intracellular iron there is to the more essential proteins.

The three sRNAs more intimately studied in this thesis are RyhB, SgrS, and ChiX, each responsible for three different signal responses: iron levels, glucose-phosphate levels, and chitin availability respectively (Bobrovskyy & Vanderpool, 2014; Chareyre & Mandin, 2018; Suzuki et al, 2015). RyhB is the key iron regulator in many bacteria, and thus is essential for cellular viability (Chareyre & Mandin, 2018). Iron is extraordinarily important for cell survival, yet mismanagement of the cellular iron levels can lead to death (Figure 13) (Oglesby-Sherrouse & Murphy, 2013). Fe^{3+} is the most abundant form of iron in the environment, but it is biologically inaccessible. All living organisms require the reduced form of iron which is Fe^{2+}. To reduce

Fe^{3+} to a biologically relevant state, many bacteria produce siderophores. However, often times biologically active iron is limited as competing organisms produce their own iron chelators to sequester iron from other organisms nearby (Oglesby-Sherrouse & Murphy, 2013). Thus, iron homeostasis must be sensitive, effective, and fast acting. RyhB is at the heart of iron metabolism in bacteria, as it regulates a host of genes all related to iron utilization in some way (Massé et al, 2005). RyhB is a 90 nt RNA controlled by its own operator which is repressed by Fur, the key iron regulating transcription factor (Massé & Gottesman, 2002). When bound to Fe^{2+}, Fur is active and represses RyhB expression. When Fe^{2+} levels are low, Fur is inactive and RyhB is expressed. RyhB negatively regulates a variety of targets mRNA which encode proteins that require Fe^{2+} for their functions, such as *sodB, sdh, sdhCDAB,* and many more (Wang et al, 2015). Through binding to Hfq and RNase E, RyhB targets these mRNAs for degradation. Simultaneously, RyhB also upregulates synthesis of proteins related to motility, iron uptake, and membrane proteins to extract iron from the environment. As RyhB requires both Hfq and the RNA degradosome to function (Deng et al, 2014; Morita & Aiba, 2011), it is an optimal target for investigation to understand the channels of communication between sRNA, Hfq, and the degradosome. In addition, RyhB controls the degradation and translation of more than 30 different mRNAs in multiple modes (Wang et al, 2015), making it an important target for understanding the intricacies of post-transcriptional regulation.

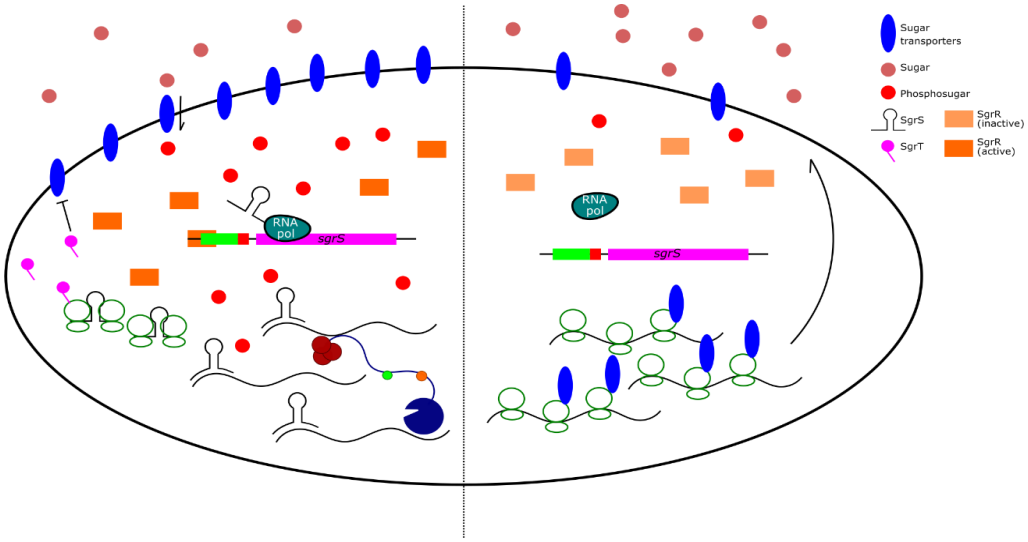


Figure 14: Simplified schematic of phosphosugar stress response mediated by SgrS. On the left hand side, there is an abundance of intracellular phosphate sugars, which activates SgrR. SgrR activates transcription of SgrS. SgrS can act in two ways, as a sRNA and when translated into a protein. The translated protein SgrT inhibits glucose and mannose transporters, while the sRNA blocks initiation of translation of the transporters, leading to the degradation of the mRNAs. Not pictured is yigL, which is stabilized by sRNA binding and requires SgrS to initiate translation. YigL dephosphorylates monosaccharides to relieve phosphosugar stress. In the absence of stress, SgrR is inactive, therefore SgrS is not produced. Sugar transporters can then be translated and transposed onto the membrane to import more sugars.

In addition to iron stress, sRNAs also control other important processes such as glucose-phosphate response. In particular, the major effector of the glucose-phosphate response is SgrS for enterobacteria (Figure 14) (Bobrovskyy & Vanderpool, 2014). SgrS is a 227 nt long sRNA which is produced during glucose-phosphate stress (Vanderpool and Gottesman, 2007). Glucose-phosphate stress in bacteria is characterized by accumulation of phosphate sugars such as glucose-6-phosphate. SgrS's primary function is to downregulate uptake of glucose and promote growth and energy metabolism, thus using up and eliminating the phosphate sugars by targeting the mRNAs of various glucose transporters and other metabolic enzymes (Papenfort et al, 2013). Currently, the described targets of SgrS include *ptsG*, the mRNA required for

glucose transporter PTS, *manXYZ*, a nonspecific glucose transporter, and *yigL*, a phosphatase (Bobrovskyy & Vanderpool, 2014). While *ptsG* and *manXYZ* are downregulated, *yigL* is upregulated. Interestingly, the sRNA SgrS can also encode a small protein called SgrT, which inhibits PtsG to block glucose import (Lloyd et al, 2017). SgrS acts upon *manXYZ* and *ptsG* in canonical ways, through base pairing with Hfq and recruitment of RNase E for endonucleolytic cleavage, then complete degradation with the degradosome (Rice & Vanderpool, 2011). However, when base paired with *yigL*, SgrS has a stabilizing function (Papenfort et al, 2013). The *yig* operon undergoes dicistron processing, after which *yigL* becomes destabilized. SgrS and Hfq binding to *yigL* obscures a second RNase E cleavage site, which protects the mRNA from complete degradation and allows YigL to be translated. The complex web of regulation controlled by SgrS showcases the important dynamics of Hfq, sRNAs, RNase E, and the RNA degradosome and how communication between the three elements is essential for life.

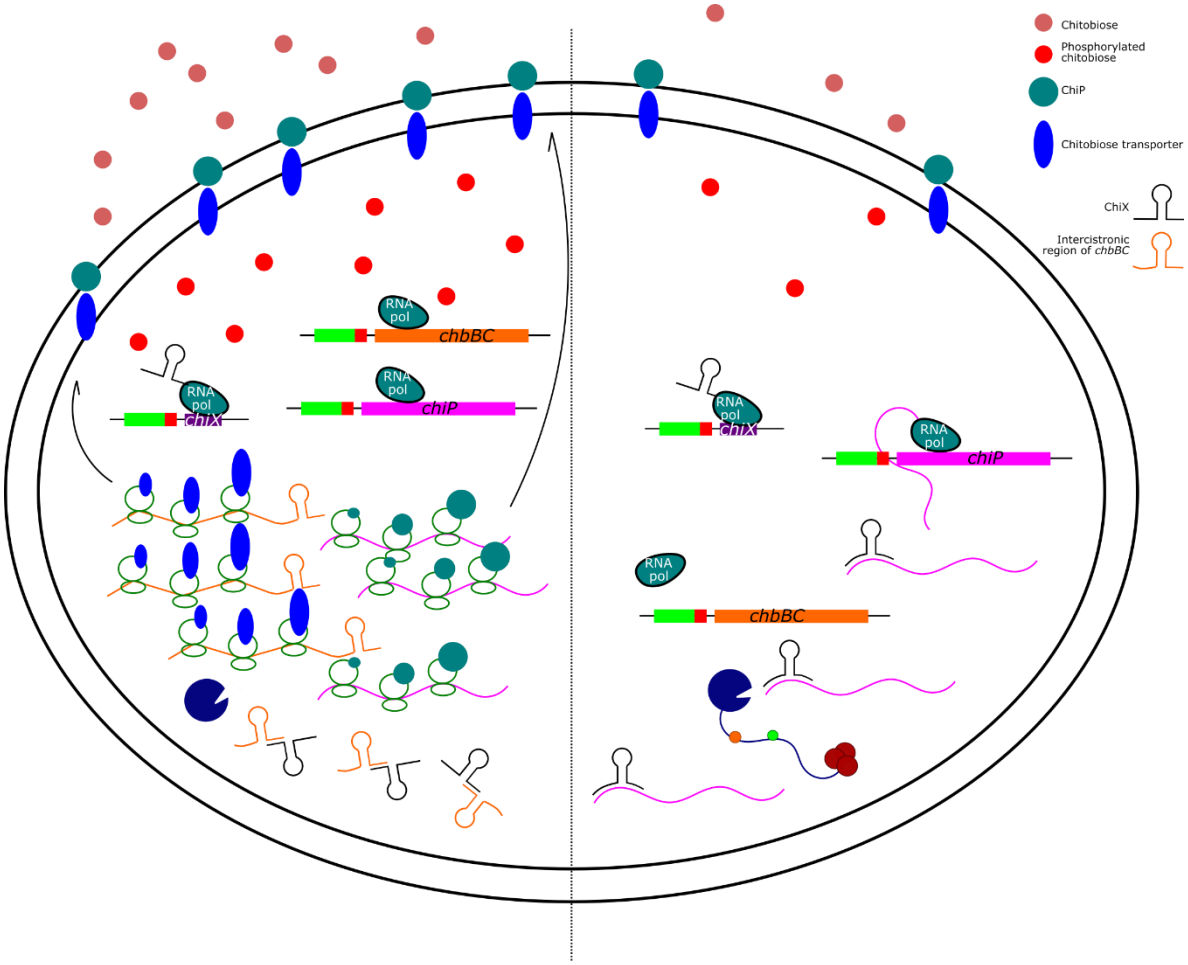


Figure 15: Schematic of chitosugar intake regulation involving ChiX. On the right, where there is not an abundance of chitobiose, ChiX and chiPQ are both constitutively expressed. However, ChiX binds the *chiPQ* translation initiation site and targets the mRNA for degradation, leading to little or no production of ChiP. On the left hand side there is an abundance of chitobiose inside and outside of the cell membrane, leading to phosphorylation of ChbR (not pictured) which activates transcription of *chbBC*. *chbBC* is translated to make the chitobiose transporter which sits on the inner membrane. The intercistronic region of chbBC also functions as an sRNA sponge, specifically base pairing to ChiX to target both sRNAs for degradation. Without ChiX, *chiPQ* is also allowed to be translated and no longer degraded, leading to an increased uptake of chitobiose.

One structurally interesting sRNA is ChiX, a class II sRNA which can bind to both faces of Hfq (Schu et al, 2015). Unlike SgrS and RyhB, ChiX serves a much more specific role in mediating post-transcriptional regulation (Figure 15) (Suzuki et al, 2015). Chitosugar metabolism is controlled by two operons, the *chiPQ* and *chbBCARFG* operons (Plumbridge et al, 2014). *chiPQ* encodes the chitoporin ChiP which is required for cellular intake of chitosugars. *chbBCARFG* encodes both a PTS transporter (*chbBCA)* and a series of metabolic enzymes used to extract energy from chitobiose (*chbRFG)*. ChiX, with the aid of Hfq, binds to the Shine-Dalgarno sequence of the *chiPQ* operon to inhibit translation initiation and begin degradation by recruitment of RNase E and the RNA degradosome. In addition, ChiX also binds to the mRNA *chbBCARFG* after *chbB,* which silences translation. However, when *chbBCARFG* expression is upregulated by release from transcription factors such as NagC, the mRNA is created then degraded (Figueroa-Bossi et al, 2009). Degradation of *chbBCARFG* creates a smaller segment of RNA which sequesters intracellular ChiX. Upon ChiX binding, the mRNA fragment-sRNA duplex is then subjected to degradation through the degradosome, with ChiX being degraded as well. Thus, the amount of cellular ChiX is lowered and repression of *chiPQ* is relieved. The cell can then intake and metabolize chitosugars as a new energy source. Recent studies have suggested that class I sRNAs like SgrS and RyhB serve more as emergency responders while class II sRNAs like ChiX serve as silencers due to a difference in binding affinity (Schu et al, 2015). Where class I sRNAs may have many binding partners, class II sRNAs have fewer binding partners but may bind in a more robust way, and could potentially displace class I sRNAs from Hfq. Nevertheless, all classes of sRNAs utilize Hfq and the degradosome for their function to a certain degree. Thus, understanding the interactions

between these three elements is essential for understanding the broader scope of sRNA mediated post-transcriptional regulation.

During its life cycle, a bacterium can experience many different stresses from the environment, most of which require a rapid response time. Mediating responses at the transcriptional level can take on the scale of hours, whereas mediation of post-transcriptional responses allows cells to adjust to the environment in minutes. For example, after induction of β-galactosidase expression by IPTG, maximum levels of β-galactosidase activity is reached in three hours (Marbach & Bettenbrock, 2012). Thus, RNA metabolism is essential for bacterial survival and proliferation. Two of the main players of post-transcriptional regulation and RNA metabolism are the multi-enzyme RNA degradosome and Hfq. With the help of various sRNAs, the RNA degradosome and Hfq can change the cellular populations of mRNAs in a rapid and controlled manner. Though a lot of research has been performed on these proteins, there still remains binding and structural properties which have yet to be described. The aim of this thesis is to further investigate the relationships between the RNA degradosome, Hfq, and sRNAs. Specifically, the structure and function of Hfq bound to ChiX and the ternary complex was studied, as well as the broader function of RhlB in the context of the degradosome through the subassembly of the ternary complex.

2 RhlB Activity impacts *E. coli* Growth

2.1 Introduction

To gain insight into the cellular functions of the RNA helicase RhlB, I used a loss of function point mutation in the DEAD-box of the helicase core to study its broad impact on cellular physiology and stress response. The D-E-A-D motif is essential for ATP hydrolysis, and point mutations in the motif have been shown to abolish ATPase activity for numerous members of the DEAD-box helicase family (Pause & Sonenberg, 1992). More specifically, swapping the glutamate to glutamine in the DEAD-box renders helicases non-functional. A close homologue of RhlB, the Vasa helicase from *Bombyx mori*, was inactivated with the DEAD -> DQAD substitution which stopped the release of hydrolysed ATP and consequently trapped the enzyme in a closed conformation while bound to RNA (Xiol et al, 2014). Using this mutation, the authors not only managed to crystallize RNA bound Vasa but also discovered transient protein-RNA complexes in vivo involving the DEAD-box helicase. I hoped to use the same mutation to both inactivate the RhlB helicase catalytic activity and trap it in a complex with RNA substrates.

To investigate the *in vivo* function of RhlB, chromosomal mutants at the DEAD box of the enzyme were generated and the impact on *E. coli* cell growth assessed. The mutant cells were grown alongside isogenic WT cells and growth was measured by recording light absorbance at 600 nm (which is related to particle scattering). As RhlB is an integral component of the RNA degradosome, mutations in the helicase are expected to potentially change degradosome activity (Khemici & Carpousis, 2004). The degradosome is an important mediator

of post-transcriptional regulation that enforces sRNA mediated degradation of target mRNAs (Bandyra et al, 2012), as well as processively degrading mRNAs destined for recycling. RhlB's role in the degradosome is to aid in the processive degradation of mRNAs by PNPase (Khemici & Carpousis, 2004). RNAs targeted for degradation are first cleaved endonucleolytically in a hydrolytic process by the N-terminal domain of RNase E, then passed onto the C-terminal associated PNPase for processive phosphorolytic degradation. As RNAs are structurally complex molecules, they may often contain higher-order secondary and tertiary folds which are difficult to degrade. Thus, RhlB plays an important role in RNA degradation by unwinding structured segments of RNA, enabling PNPase to function smoothly (Khemici & Carpousis, 2004). Although RNA degradation is important during regular growth, it is potentially more important during periods of stress. The RNA degradosome regulates the half-life of certain mRNAs (Bobrovskyy & Vanderpool, 2016; Battesti et al, 2011; Massé & Gottesman, 2002), and a functional degradosome is integral for a rapid stress response. Hence, the role of RhlB may be more prominently displayed during periods of stress, as disruption to RNA degradation may be detrimental to survival under challenging conditions.

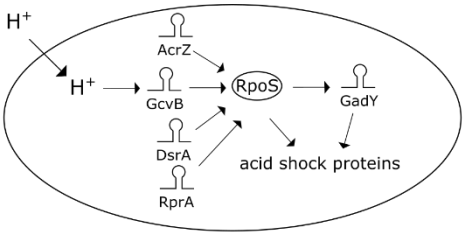


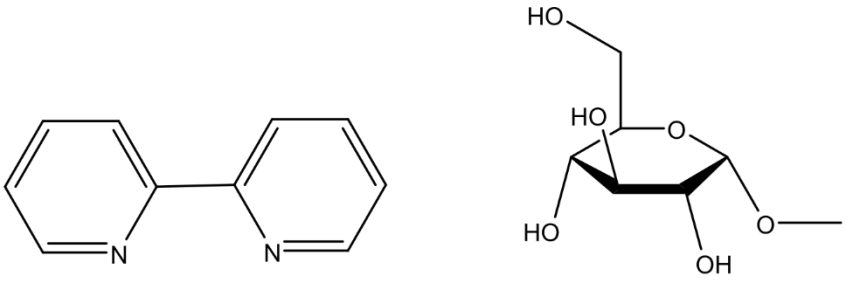
Figure 16: *E. coli* acid shock response. In the presence of high levels of acid, certain sRNAs such as AcrZ, GcvB, DsrA, and RprA are expressed, inducing RpoS expression (Jin et al, 2009). Increased amount of RpoS leads to increased expression of sRNA GadY (Opdyke et al, 2004) as well as various acid shock proteins such as protein chaperones (Zhao and Houry, 2010).

The RhlB DQAD mutant cells were thus subjected to different kinds of stress to assess their response in the absence of RhlB catalytic activity. Three different stress conditions — acid stress, iron stress, and phospho-sugar stress — were selected as the RNA degradosome was known to be involved in the stress response of the aforementioned conditions (Chung et al, 2006; Battesti et al, 2011; Deng et al, 2014; Morita & Aiba, 2011; Rice & Vanderpool, 2011). When *E. coli* cells are grown in a low pH of around 5, RpoS is activated as a general transcription factor by sRNAs (Chung et al, 2006) (Figure 16). Under normal growth conditions, the RpoS levels in the cell are very low (Battesti et al, 2011), as RpoS mRNA is very rarely translated and any protein product of the transcript is rapidly degraded. The start codon of RpoS is buried under extensive secondary structure (Battesti et al, 2011) which prevents ribosomal binding and translation initiation. Thus, the bulk of RpoS mRNA is unprotected and rapidly degraded by the RNA degradosome. However, under stress conditions, translation of RpoS becomes active (Battesti et al, 2011). sRNAs such as ArcZ, DsrA, and RprA are transcribed, which bind to the 5' end of RpoS, releasing the ribosome binding site and allowing translation to begin. Translating ribosomes make RpoS protein while simultaneously protecting RpoS mRNA from degradation (Battesti et al, 2011). As such, *E. coli*'s response to low pH involves the activity of sRNAs and the functional degradosome to support active degradation which is required for RpoS function. Hence, if RhlB activity is required for RpoS degradation, the cellular response to acid stress could be disabled or differently executed by the RhlB DQAD mutant. Similarly, under iron

stress and phospho-sugar stress conditions, both sRNAs and the RNA degradosome are involved in stress mediation. By subjecting the mutant and WT cells to stress, I hoped to understand under which conditions RhlB is necessary for a rapid cellular response and to what extent RhlB is utilized in the stress response.

To induce stress artificially, two chemicals were used along with pH-controlled LB media (Figure 17). 2,2-bipyridyl is a high affinity chelator of iron, sequestering it from use in bacteria. With the addition of 2,2-bipyridyl to the media, cells will experience a depletion of biologically available iron which will trigger a measurable stress response. Secondly, α-D-methylglucoside was used to induce phospho-sugar stress response. α-D-methylglucoside is a sugar mimic that, once phosphorylated, cannot be exported and therefore accumulates in the cell. When a high level of phosphorylated sugars builds in the cell, SgrS is activated and signals for degradation of many different mRNAs such as *folE* and *adiX* (Bobrovskyy & Vanderpool, 2016). Removal of the RNase E C-terminal domain seems to inhibit SgrS regulation of the two RNAs (Bobrovskyy & Vanderpool, 2016). Increased expression of *folE* and *adiY* — encoding GTP cyclohydrolase I and a transcription factor regulating arginine decarboxylase genes respectively — was also shown to be detrimental to cell growth during phosphor-sugar stress (Bobrovskyy & Vanderpool, 2016). Thus, we believe that inducing the phosphor-sugar stress response in the cell may influence the growth of RhlB DQAD cell lines differently compared to WT cell lines, as the response to stress implicates C-terminal constitutive degradation of the degradosome. Although bacteria can respond to stress quite rapidly (Massé & Gottesman, 2002), I wanted to explore first if the inactivation of RhlB may result in a prolonged effect on global physiology. Therefore this chapter will examine if inactivation of RhlB creates a

measurable impact on the growth of mutant cells compared to WT cells under the aforementioned stress conditions.



2,2-bipyridyl α-D-methylglucoside

Figure 17: Chemical structures of 2,2-bipyridyl and α-D-methylglucoside. Three molecules of 2,2-bipyridyl sequester Fe^{2+} using the nitrogens to chelate free iron, thereby removing soluble iron ions from solution. Decreased iron concentrations will thereby trigger the iron-stress response. α-D-methylglucoside can be imported into the cell and phosphorylated, but cannot be dephosphorylated. Phosphorylated sugars also cannot be exported. Thus, phosphorylated α-D-methylglucoside will accumulate in the cell triggering the phospho-sugar stress response.

2.2 Results

To explore the consequences of RhlB dysfunction in live cells, we constructed, with collaborators in Toulouse (Carpousis laboratory), two strains of *Escherichia coli* with chromosomal mutations using the lambda red recombination system (Datsenko and Wanner, 2000). The mutant strains contained a point mutation at nucleotide 498 of *rhlB*, creating the RhlB E166Q genotype. A kanR cassette was inserted after the RhlB E166Q gene to be co-transcribed, and either a streptavidin binding tag or a hemagglutinin tag was added to the N-terminus of RNase E (Figure 18 and Table 1). The affinity tags were introduced into the RNase

E gene with the plan of attempting pull downs for future analysis of the bound RNA once optimal growth conditions of mutant and parental strains could be identified. To evaluate growth characteristics in liquid culture, the OD_{600} was measured every 30 minutes until the readings reached around 1 unit, or cells showed signs of dying with decreasing OD.

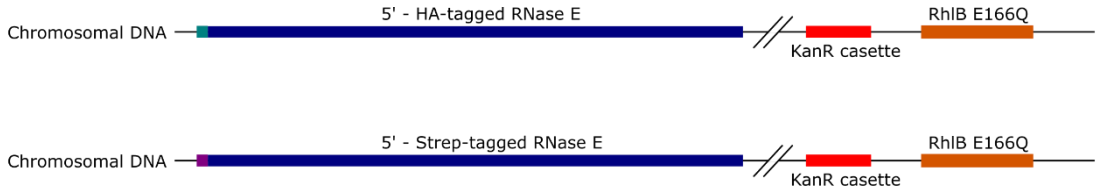


Figure 18: The chromosomal organization of the mutants. Both mutants contain a kanamycin resistance cassette which allows isogenic strains to be grown under less stringent sterility conditions. Both strains contain a point mutation which translates into a RhlB protein with a glutamine at position 166 instead of glutamate (E166Q). The indicated tags, hemagluttinin (HA) and streptavidin (strep), are fused to the N-terminal of RNase E.

Table 1. List of strains used in this study

Background	**NCM3416 (Soupene et al, 2003).**
SAJ103	rhlB (DQAD)-kan. This kan marker is co-transcribed with RhlB and cannot be removed.
SAJ104	rne-STREP, rhlB (DQAD)-kan.
SAJ122	rne-HA, rhlB (DQAD)-kan
WT	K12 (MG1655)

SAJ104 and 122 were made by phage P1 transduction.

Each cell line was grown under five conditions including a control (grown in LB media only). The conditions of interest tested were acid stress, iron stress, combined acid and iron stress, and combined acid and phospho-sugar stress (Figure 19). Unexpectedly, both mutants seemed to grow faster than the WT cells under all stress conditions. In iron stress conditions, the discrepancy between the growth rates seem to be diminished when compared to the discrepancy between the growth rates in control conditions (Figure 19B, Table 2). In conditions of acid stress, no change in growth discrepancy was noted. With compound conditions (i.e. growth in the presence of multiple stressors) all *E. coli* strains, both mutants and WT, had significantly slower rates of growth (Table 2). In the case of both acid and iron stress, all cell lines seem to experience a decline in population during the log phase, with HA tagged mutant cells having the most dramatic population drop (Figure 19D). In the presence of both acid stress and phospho-sugar stress, the difference in growth between mutant cell lines and the WT cell line seems to be exacerbated (Figure 19E) compared to the difference in growth rate of the cells when grown in LB.

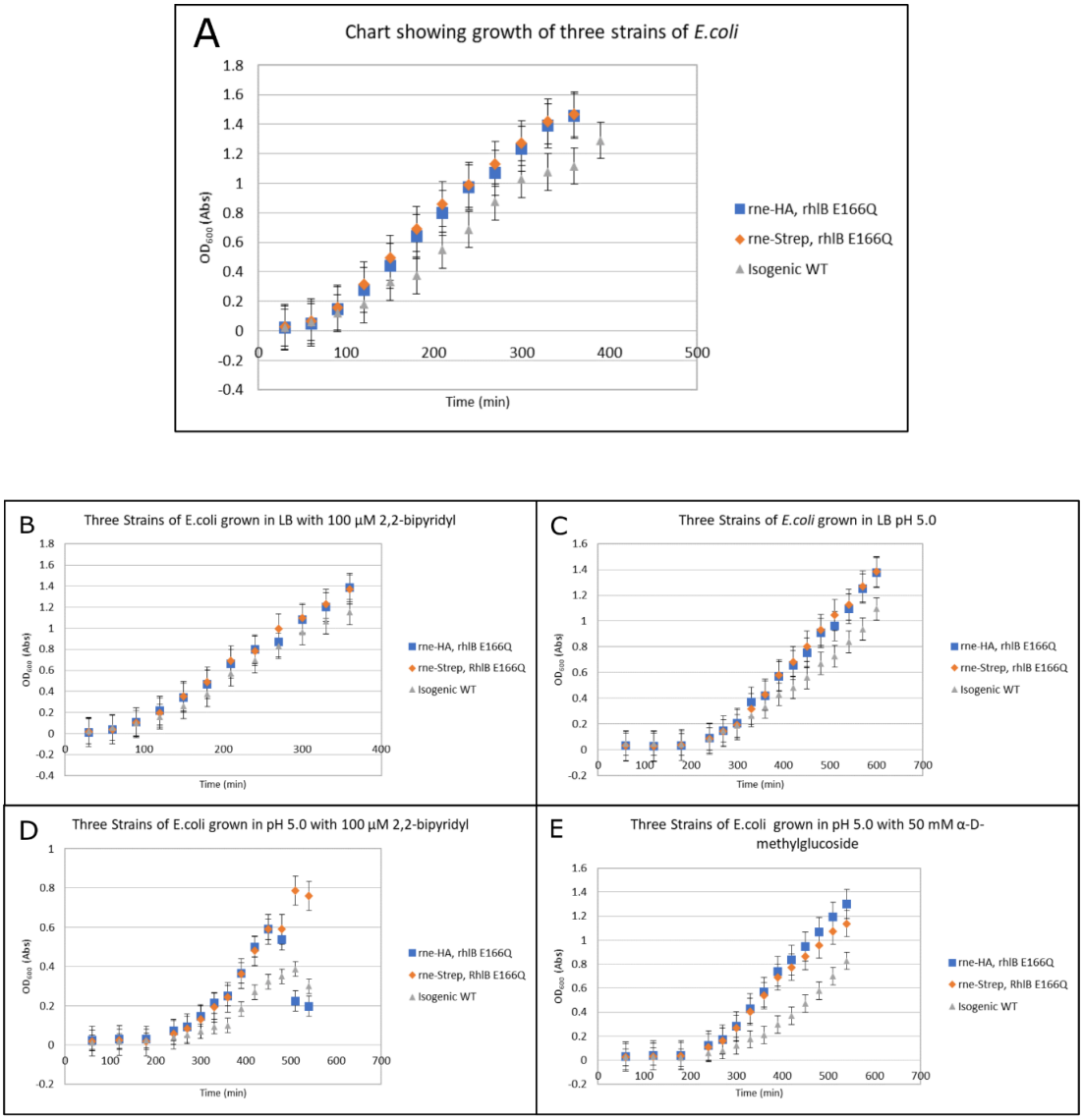


Figure 19: Graphs comparing the growth profiles of three different strains of *E. coli*. Datapoints show the OD600 measured at each time. All data was measured in triplicate, the standard error is displayed in the error bars. In graph A, all strains were grown in LB media. The different tags show little to no effect on the growth of the bacteria, but the mutant strains are shown to grow faster than the WT strain. In graph B, all strains were grown in LB with 100 µM of 2,2-bipyridyl, an iron chelator. Under iron stress conditions, the difference in growth speed seems to be minimized. In graph C, all strains were grown in LB media at pH 5.0. Under acid stress, the difference in growth speed between the mutant strains and the WT strain seems to be accentuated. In graph D, all strains were exposed to both acid stress and iron stress. WT cells seem to grow slower, but experienced a lesser decline in OD compared to the HA tagged mutant strain. However, the Strep-tagged strain only seemed to decline in growth slightly. In graph E, all strains were exposed to both acid stress and phospho-sugar stress. The difference in growth between mutant strains and the WT strain seems more pronounced than only acid stress, suggesting a potential additive effect between phospho-sugar and acid stress on the slowed growth of WT *E.coli*. Calculated growth rates are presented in Table 2.

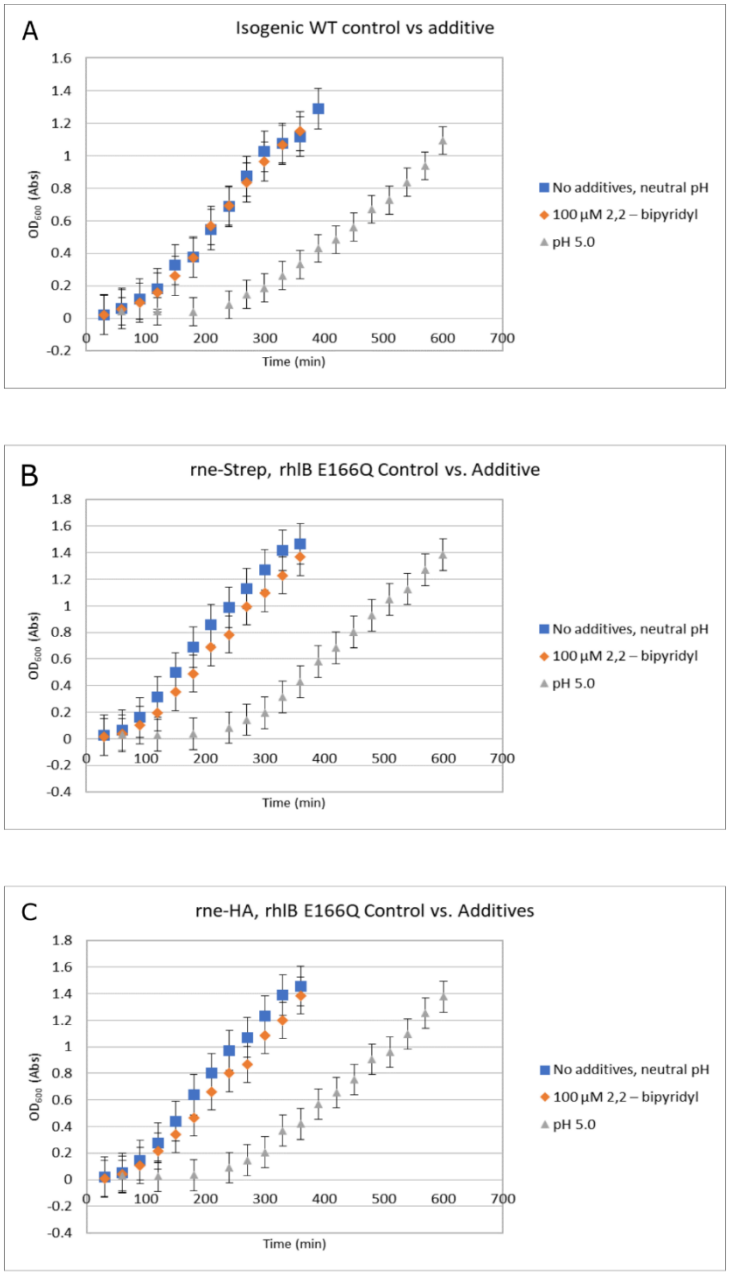


Figure 20: Growth curves showing the effect of each additive on individual strains. All datapoints were measured using a UV spectrophotometer at either 30 minute (all conditions) or 60 minute (in the beginning of pH 5.0 measurements) intervals. All cells were grown in triplicate. The datapoints plotted are the mean of each timepoint, and the error bars show the stardard error calculated from all data points in a single series (i.e. WT cells, no additives). In all graphs, pH seems to have the largest negative effect on cell growth rate. Interestingly, 100 µM of 2,2-bipyridyl is shown to have a slight effect only on the growth of the two mutant strains. In contrast, 2,2,-bipyridyl has no discernable effect on the growth of the WT strain. Calculated doubling times are presented in Table 2.

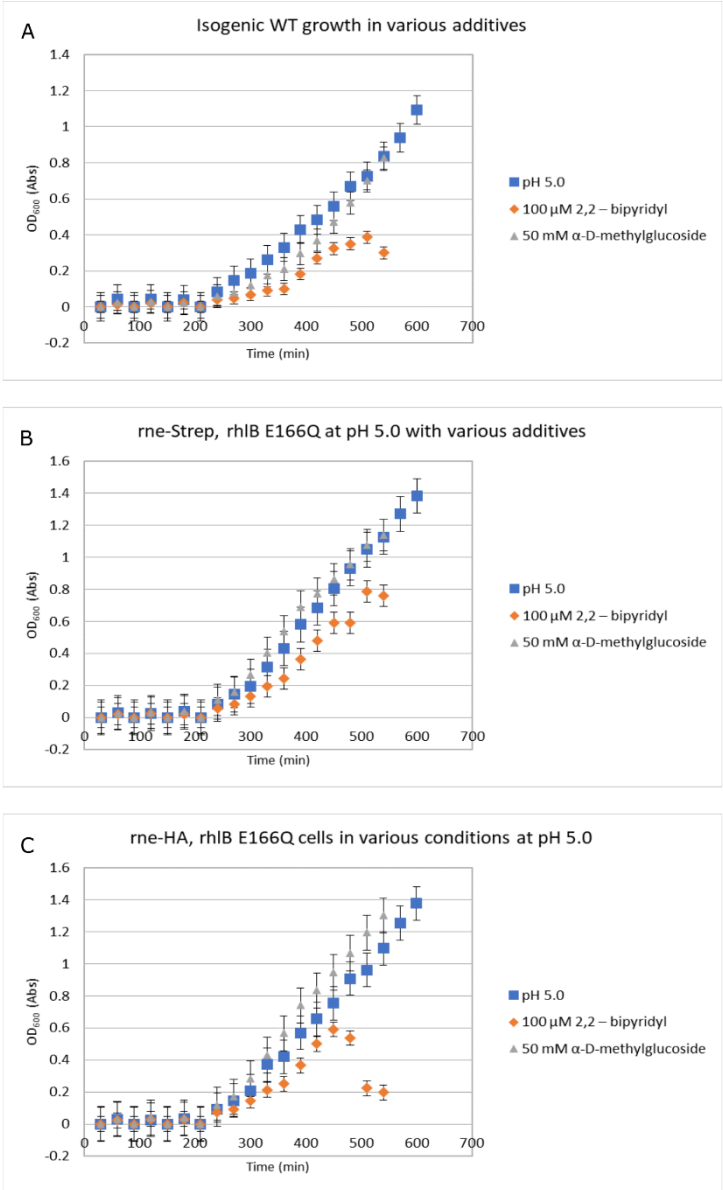


Figure 21: Graphs showing the effect of compound stress on different strains of *E. coli*. All strains were grown in pH 5.0 (shown in blue), pH 5.0 plus 2,2-bipyridyl (shown in orange), and pH 5.0 plus 50 mM α-D-methylglucoside (shown in grey). Graph A shows the WT response to compound stress, while graph B shows the response of Strep-tagged RNase E and RhlB-E166Q cells to compound stress. Similarly, graph C also shows the response of cells that express the RhlB-E166Q mutation, although in this case RNase E is HA tagged instead. The graphs show that the growth of all strains in the presence of acid stress and iron stress is inviable. When both phospho-sugar stress and pH stress are combined, cells react in different ways depending on the strain. In the WT strain (graph A), the presence of both stressors grow slightly slower than the presence of only pH stress. However, in mutant Strep-tagged cells (graph B), the presence of both stressors has no effect compared to the presence of only pH stress. In HA-tagged cells (graph C), the presence of both stressors seemed to have a positive effect on the cell growth compared to the presence of only pH stress. However, note that although a slight difference in growth pattern is observed, the error bars of each condition overlap with each other slightly during the most notable growth periods, indicating that a larger sample size may be needed for more certainty.

Comparing the relationship between different kinds of stress placed upon a single cell strain, a different picture emerges (Figure 20). The stress that delays cell growth the most is acid stress in all cell lines. Note that while 2,2-bipyridyl has no effect on the mutant cells, it has a slight positive effect on WT cells (Figure 20 and Table 2). Though a slight difference can be seen, note that the error bars of no additives versus 2,2-bipyridyl plots in both Figure 20B and Figure 20C overlap, suggesting that more data is needed to make a definitive statement about whether the difference is meaningful.

Figure 21 shows the effect of compound stressors on the growth of mutant and WT cells. Each subplot shows the growth of one cell line under three different conditions. The three conditions plotted are acid stress, acid plus iron stress, and acid plus phosphor-sugar stress. In all plots, combined acid and iron depletion stress was the most debilitating condition for cell growth. All cell strains showed population decline towards the end of the plot. With combined phosphor-sugar and acid stresses, the three strains respond differently. WT cells showed slowed growth whereas both mutant cell strains showed slightly faster growth under both stressors present compared to only acid stress. Like the plot above, there are overlaps in the error bars, suggesting that further investigations need to be done for a definitive conclusion.

Table 2: Lag time, maximum doubling time, entry into stationary phase/death phase of each strain in all conditions (min).

Cell line / **Additive**	**Isogenic WT**		**rne-HA, rhlB(E166Q)**		**rne-Strep, rhlB(E166Q)**	
No additives	Lag time:	N/A	Lag time:	N/A	Lag time:	N/A
	Max doubling time:	91.2	Max doubling time:	55.5	Max doubling time:	63.2
	Stationary/death phase entry time:	N/A	Stationary/death phase entry time:	N/A	Stationary/death phase entry time:	N/A
2,2-bipyridyl	Lag time:	N/A	Lag time:	N/A	Lag time:	N/A
	Max doubling time:	49.1	Max doubling time:	59.1	Max doubling time:	60.8
	Stationary/death phase entry time:	N/A	Stationary/death phase entry time:	N/A	Stationary/death phase entry time:	N/A
pH 5.0	Lag time:	180	Lag time:	180	Lag time:	180
	Max doubling time:	159.6	Max doubling time:	146.9	Max doubling time:	126.8

	Stationary/death phase entry time:	N/A	Stationary/death phase entry time:	N/A	Stationary/death phase entry time:	N/A
2,2-bipyridyl + pH 5.0	Lag time:	180	Lag time:	180	Lag time:	180
	Max doubling time:	61.8	Max doubling time:	46.5	Max doubling time:	53.9
	Stationary/death phase entry time:	510	Stationary/death phase entry time:	450	Stationary/death phase entry time:	N/A
α-D-methylglucoside + pH 5.0	Lag time:	180	Lag time:	180	Lag time:	180
	Max doubling time:	107.3	Max doubling time:	77.9	Max doubling time:	82.7
	Stationary/death phase entry time:	N/A	Stationary/death phase entry time:	N/A	Stationary/death phase entry time:	N/A

Though growth curve plots provide an accessible visual to the health of the cells under various growth conditions, descriptions of lag time, max doubling time, and entry into stationary/death phase time (Table 2) are more apt descriptors of overall population health. Lag time was calculated as the last timepoint in which no increase in OD600 was measured, while entry into stationary/death phase was calculated as the first timepoint in which OD600 fell below the previous timepoint. Max doubling time was calculated using the formula $D_t = t/n$, and $N_t=N_0(2)^n$, where D_t = doubling time, t= time, n=number of generations (2), N_t = OD600 after t has passed, and N_0 = initial OD600. D_t was calculated during log phase (OD of 0.6), and thus N_0 and N_t were taken at the time points immediately before and after OD crossed 0.6.

In most of the reported growth curves, data was collected from the log phase, and thus only maximum doubling time is presented. In Table 2, maximum doubling time of rne-HA or rne-Strep tagged rhlB E166Q are usually within 10 minutes of each other, with the exception of pH 5.0. However, the maximum doubling rate of isogenic WT is much higher than either

tagged mutant strain, ranging from around 10 mins (pH 5.0 + 2,2-bipyridy) to around 30 mins (no additives) longer. One notable exception is the 2,2-bipyridyl condition, where the WT strain grew at a faster rate than the mutant strains. A lag time in bacterial growth only appeared in acid stressed conditions. However, increasing the amount of stress or differences in the strains did not change the lag time. For most conditions, entry into stationary/death phase was not recorded. Entry into stationary/death phase was only recorded in the compound condition of acid stress and iron stress. Both the WT and the rne-HA mutant were affected. WT cells entered stationary/death phase after 510 minutes of inoculation, while rne-HA mutant entered stationary/death phase after 450 minutes. Though rne-Strep tagged mutant did not enter stationary/death phase conclusively, there was a marked stall in cellular growth at 450 minutes after inoculation. These data taken together suggest that, generally speaking, RhlB E166Q mutants grow faster than WT cells. Acidic conditions delay the cell's entry into log phase, and acidic conditions plus iron depletion are the most detrimental to cell growth, with WT cells being slightly less susceptible to population death than mutant cells despite mutant cells having a faster growth.

2.3 Discussion

Though mutant strains and the WT strain display broadly similar growth profiles, small differences in growth rate can be seen in the results. The WT cells, broadly speaking, grow slower than the mutant cells. However, the mutant loses its capacity to grow faster than the parental strain under iron-starvation conditions, where WT cells seem to grow at an accelerated rate induced by the removal of iron, whereas mutant cells are unaffected. All cells grow at a

much slower rate in LB with a pH of 5.0, but with the addition of α-D-methylglucoside, mutant cells seem to increase in growth rate while WT cells decrease (Table 2). The pertinent question in this case is to ask whether these results can be attributed to the mutation generated on RhlB or whether these results arose from secondary effects due to the difference in genetic background. Several differences in the genetic background can be found, some resulting from engineering and others not. The understanding of the effects of these background changes may elucidate the overall effect that the RhlB E166Q point mutation has in the cell.

Firstly, *E. coli* MG1655 and NCM3416 are different. Previously, *E. coli* MG1655 has been shown to have certain growth deficiencies when cultured in minimal media. MG1655 shows growth deficiencies such as partial pyrimidine requirement, slow growth on galactose, and slow growth on glycerol (Soupene et al, 2003). *E. coli* NCM3416, on the other hand, has been corrected for these growth defects and therefore should not display any growth defects (Soupene et al, 2003). More recently, this difference in genetic makeup has been shown to cause a slower growth in the MG1655 strains as compared to the NCM3416 strains (Hadjeras et al, 2019). Thus, it is highly likely that the patterns in growth discrepancies seen in the results is a product of the different genetic makeup between the WT isogenic strains and the mutant RhlB E166Q strains rather than the point mutation itself. Interestingly, the deficiency in growth rate was abolished upon addition of 2,2-bipyridyl. One possible explanation for this increase in growth may be the pyrimidine requirement shown by MG1655 strains (Soupene et al, 2003). Upon induction of iron stress, many mRNAs are degraded to conserve iron (Wang et al, 2015). Degradation of these mRNAs may in fact provide the necessary pyrimidines for optimal cellular function, and thus increase the growth rate to match that of the mutant cell lines in the NCM3416 background.

Other genetic differences between the WT strains and the mutant strains consist of scars and tags which have been introduced to the mutant chromosome during the creation of the mutant strains. Though it is possible that the presence of the tags has caused the discrepancies shown in the growth curve, it is highly unlikely that either addition affects the growth in any significant way. The two differently tagged strains behave in a similar way as shown in the results. Additionally, RNase E has been previously tagged with many different tags on the N-terminal without disturbance to degradosome function (Hadjeras et al, 2019; Khemici et al, 2008). The scars were engineered to be created in untranslated regions, and therefore are unlikely to disrupt protein function.

Other considerations concerning the RhlB E166Q contribution to the difference in growth rate include how disruptive the mutation is to the function of the protein as well as how essential the protein is in the cell. Previous studies have shown that deletion of RhlB simply increased the population of certain structured untranslated regions of mRNAs in the cell (Khemici & Carpousis, 2004), and thus, engineering a point mutation on RhlB may cause a similar effect which may not have exerted a noticeable effect on cellular growth rate. With all considerations taken into account, the discrepancies in growth rate seen in the results are unlikely to be caused directly by the RhlB E166Q mutation. Experiments focusing on short term sRNA response and global RNA sequencing may yield more information regarding the general role of RhlB in the cell.

2.4 Materials and Methods

2.4.1 Lambda Red recombination

Lambda Red recombination was performed on *E. coli* K12 strains containing His and Strep-tagged RNase E kindly made by our collaborators in Toulouse, AJ Carpousis and colleagues. The list of primers is shown below in Table 3. PCR was performed using designed primers for *E. coli* K12 genomic DNA. PCR products were purified by gel-extraction. PCR products were transformed into pre-existing strains which contained pKD46 provided by our collaborator AJ Carpousis by electroporation. pKD46 provides the 1,894 nt (31348–33241) and 2,154 nt (31088–33241) region of bacteriophage phage λ under an arabinose promoter (Datsenko & Wanner, 2000). Transformed cells were grown in SOC at 37ºC for 2 hours to induce homologous recombination. Transformed cells were then plated onto kanamycin plates to select for recombinant cell lines. Recombinant cell lines were then streaked onto ampicillin plates to check for loss of pKD46 plasmid. Appropriate cell lines were checked for genomic mutation by sequencing.

Table 3. List of primers used in strain construction

Primer name	**Primer sequence (5' - 3')**	**Primer usage**
OAJ9s	GCGTGTTGGTTTCGTTGTTGG	Amplifies 1757 bp region containing rhlB
OAJ10s	GTAGCGACAACGCGAATTTGC	
OAJ11e	GCCGCTCCTGCGCGTCCGCAACCTGTTGAGT GGAGCCACCCGCAGTTCGAAAAATAGTAAT GTAGGCTGGAGCTGCTTCGA	

OAJ12e	TGGCAGTTACCAGGGCTTGATTACTTTGAG ATATGAATATCCTCCTTAGTTCC	Amplifies 1102 bp for rne-STREP construction
OAJ18e	CTCGGTGCCATTCAGGTGGTGGTACTGGAC CAGGCCGATCGCATGTACGATCTG	Amplifies 828 bp product from rhlB: 5' DQAD, 3' crossover
OAJ19e	CAGTAATACAAGGGGTGTTATGAG	
OAJ20s	GATACAGTTTGAATGATTTTGAGTATG	Amplifies 873 bp product from pACYC177: 5' crossover, 3' rhlB/gpp homology
OAJ21s	CTCGGTGCCATTCAGGTGGTGGTAC	Amplifies 1677 bp product from rhlB-DQAD kanR crossover
OAJ22e	GATACAGTTTGAATGATTTTGAGTATGACAT TTTTTATTTAGAAAAACTCATCGAGCATCAA ATG	Amplifies 873 bp product from pACYC177: 5' crossover, 3' rhlB/gpp homology
OAJ23s	CTCGGTGCCATTCAGGTGGTGGTAC	Amplifies 1677 bp product from rhlB-DQAD kanR crossover

2.4.2 Cell growth

WT cells were streaked onto LB-Agar plates. Mutant cells were streaked onto LB-Agar plates with 30 μg/mL of kanamycin. All plates were incubated at 37°C over 16 h. Following incubation, one colony was picked and grown in 5 mL of LB (for WT cells) or LB plus 30 μg/mL of kanamycin (for mutant cells). Inoculated starter cultures were grown at 37°C over 16 h with aeration. Starter culture was added to cultures containing the appropriate condition at a 1:100 ratio. Three cultures were grown for every condition and cell strain. Stress condition cultures were incubated at 37°C with aeration. OD_{600} was measured every 30 min after inoculation.

3 BIOPHYSICAL AND PHYSIOLOGICAL ANALYSES OF RHLB E166Q MUTANT

3.1 Introduction

Following the unsuccessful attempt to understand the cellular role of RhlB through growth rate measurements, I attempted to characterize the biochemical properties of the RhlB E166Q mutant to verify whether it significantly impacts RhlB function. In the report of Xiol et al (2014), Vasa — an insect homologue of RhlB — captured transiently associating piwi complexes with an E to Q point mutation in the Walker B motif. This simple mutation blocked the release of bound RNA from the Vasa helicase, as the substituted glutamine residue creates an extra H-bond with the inorganic phosphate liberated by ATP hydrolysis. Preventing the release of the inorganic phosphate prevents the transition to an open configuration of the two RecA domains with concomitant release of bound RNA. Using this mutant as an investigative tool, the authors were able to identify and characterize the highly transient Amplifier complex. Through engineering the same mutation onto *E. coli* RhlB, I anticipated a similar effect — clamping of the helicase onto RNA may be achieved. Currently, the full extent of RhlB's participation in the degradosome's degradative, processive, and surveillance activities has yet to be defined. As RNAs are numerous, diverse in structure and function and are expected to interact with RhlB only transiently, many RNA substrates of the helicase may remain to be discovered. A RhlB RNA clamp could be used to elucidate transient RhlB-RNA complexes which may be otherwise difficult to capture by conventional methods.

Though the analysis of the mutant *E. coli* RhlB E166Q growth in liquid culture was inconclusive, biophysical properties associated with the catalytic inactivation of the enzyme may still yield interesting characteristics. In particular, I wanted to investigate whether the E166Q mutation changes the K_d for RhlB binding to RNA, similar to the effects seen by Xiol *et al* in the mutation of the *Drosophila* Vasa helicase. Using this knowledge, I then hoped to identify RNAs which are bound to RhlB *in vivo*. To investigate how tightly RhlB bound to RNA, I chose the intergenic region between *malE* and *malF* as the RNA binding target. *malEFG* encodes three different proteins which make up the ABC transporter responsible for importing extracellular maltose (Bedouelle et al, 1982). The *malEFG* operon is under the post-transcriptional regulation by the RNA degradosome, making it a highly relevant target. Moreover, RhlB has been shown to be specifically implicated in the degradation of structured intergenic regions between *malE* and *malF* (Khemici & Carpousis, 2004). Using MalEF as the RNA binding target, I compared affinities for WT and mutant RhlB. I utilized the same mutant described in the previous chapter in an attempt to capture transiently associated RNAs on RhlB *in vivo*.

To more accurately mimic the *in vivo* activity of RhlB *in vitro*, I used a complex representing a subassembly of the degradosome that encompasses the RNA binding and helicase functions. This assembly consists of RNase E amino acids 603-850, RhlB, and enolase (denoted as the ternary complex) (Figure 9). RhlB ATPase activity has been shown to increase upon binding to the RhlB binding site on the RNase E C-terminal domain (Chandran et al, 2007). Thus, addition of RNase E 603-850 in complex with RhlB may yield a more accurate description of RhlB activity *in vivo*. Enolase was also added as it was found to potentially modulate binding of certain RNAs (Morita et al, 2004; Murashko and Lin-Chao, 2017).

Together, these three elements may more accurately represent the activity of RhlB bound to the C-terminal domain of the RNase E in the degradosome assembly.

3.2 Results

3.2.1 Expression and purification of the ternary complex (Bruce et al, 2018) of RhlB, enolase and the recognition segment of RNase E

To investigate RNA binding by RhlB within the context of the RNA degradosome, I prepared the stable complex of the helicase with enolase and the recognition segment of RNase E, encompassing residues 603 to 805 (Bruce et al, 2018) (Figure 22).

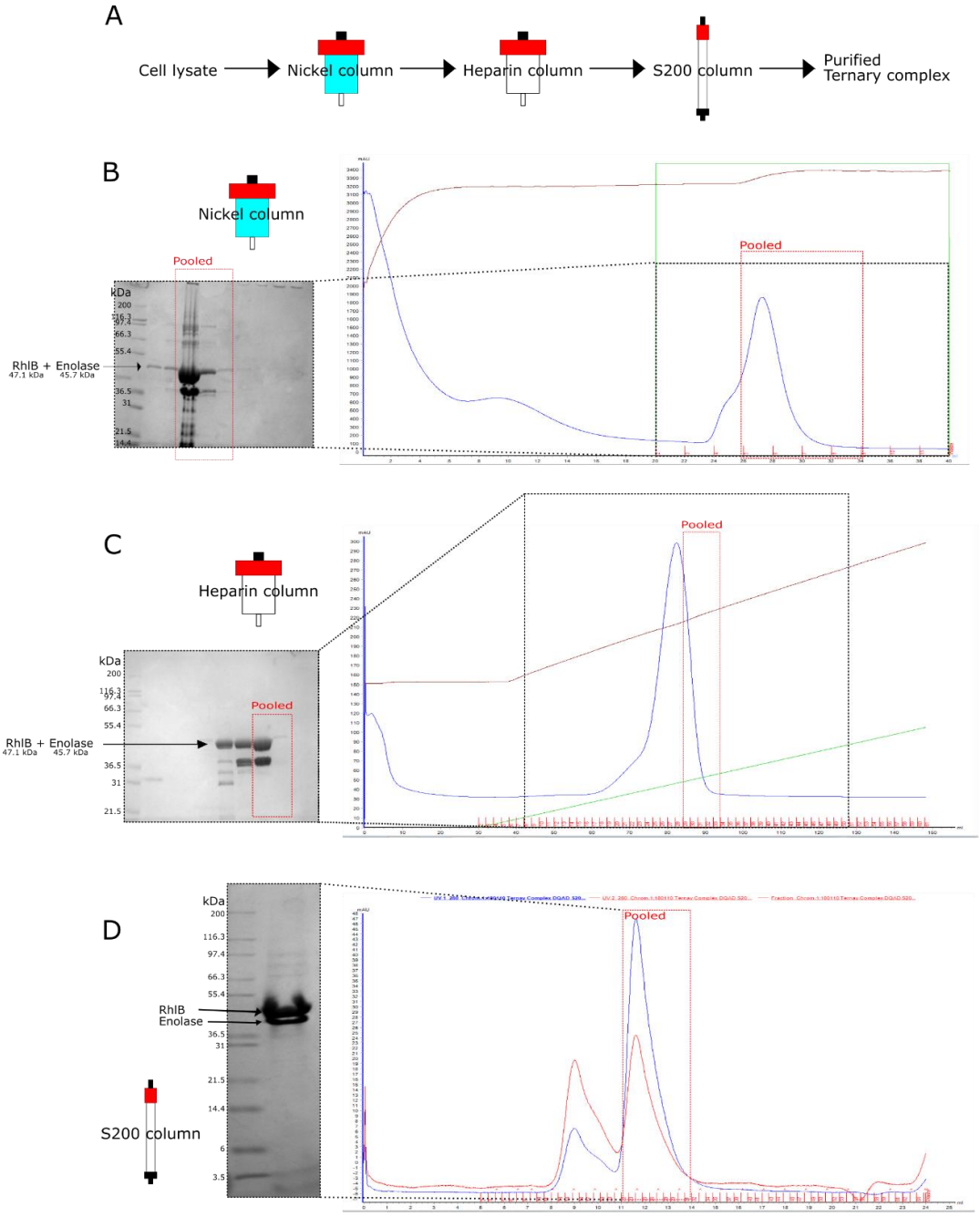


Figure 22: Purification of ternary complex. A) Schematic of purification process. Cell lysate, after clarification by centrifugation, is first loaded onto a nickel column for affinity purification through the 6xHis tag, followed by anion exchange chromatography with the heparin column, then size exclusion purification with the S200 column. B) UV absorbance trace from affinity chromatography and resulting SDS-PAGE analysis. In the chromatogram, the blue line indicates UV absorbance at 280nm, the green line indicates percentage elution buffer, and the brown line indicates conductance. The red box indicates the fractions which were pooled and further purified. The red box on the gel shows the proteins and the purities obtained. C) Chromatogram and SDS-PAGE gel of anion exchange chromatography. D) Chromatogram and SDS-PAGE gel of size exclusion chromatography.

3.2.2 RNA binding

To evaluate RhlB interactions with RNA, an EMSA was performed with MalEF at constant concentration pre-incubated with increasing concentrations of WT ternary complex and ternary complex including RhlB E166Q. The best EMSA results were obtained with agarose rather than polyacrylamide as support matrix (results not shown), and Figure 23 shows a representative agarose gel stained with SYBR gold dye to visualize RNA. Two titrations were run on one gel. The left half of the gel showed the shift of RNA with increasing amounts of WT RhlB, while the right half of the gel showed the shift of RNA with increasing amounts of RhlB E166Q. Significant bands are highlighted in a red rectangle. From the gel, binding of RhlB E166Q to MalEF is evidently tighter than WT RhlB binding. Interestingly, binding of RhlB to MalEF seems to be in a ratio of 1:1 despite MalEF being quite long (2.8 knt) as only one higher order band is seen at physiologically relevant concentrations of RhlB. However, when RhlB E166Q concentration is at 10x RNA concentration, a faint band can be seen above the dominant shifted band. These data suggest that the mutant RhlB has a preferred binding site on MalEF but will bind other sites with much lower affinity. As the affinity of WT RhlB for MalEF is much lower than that of the mutant, it is unlikely that the binding site can be saturated by the native enzyme in the absence of ATP.

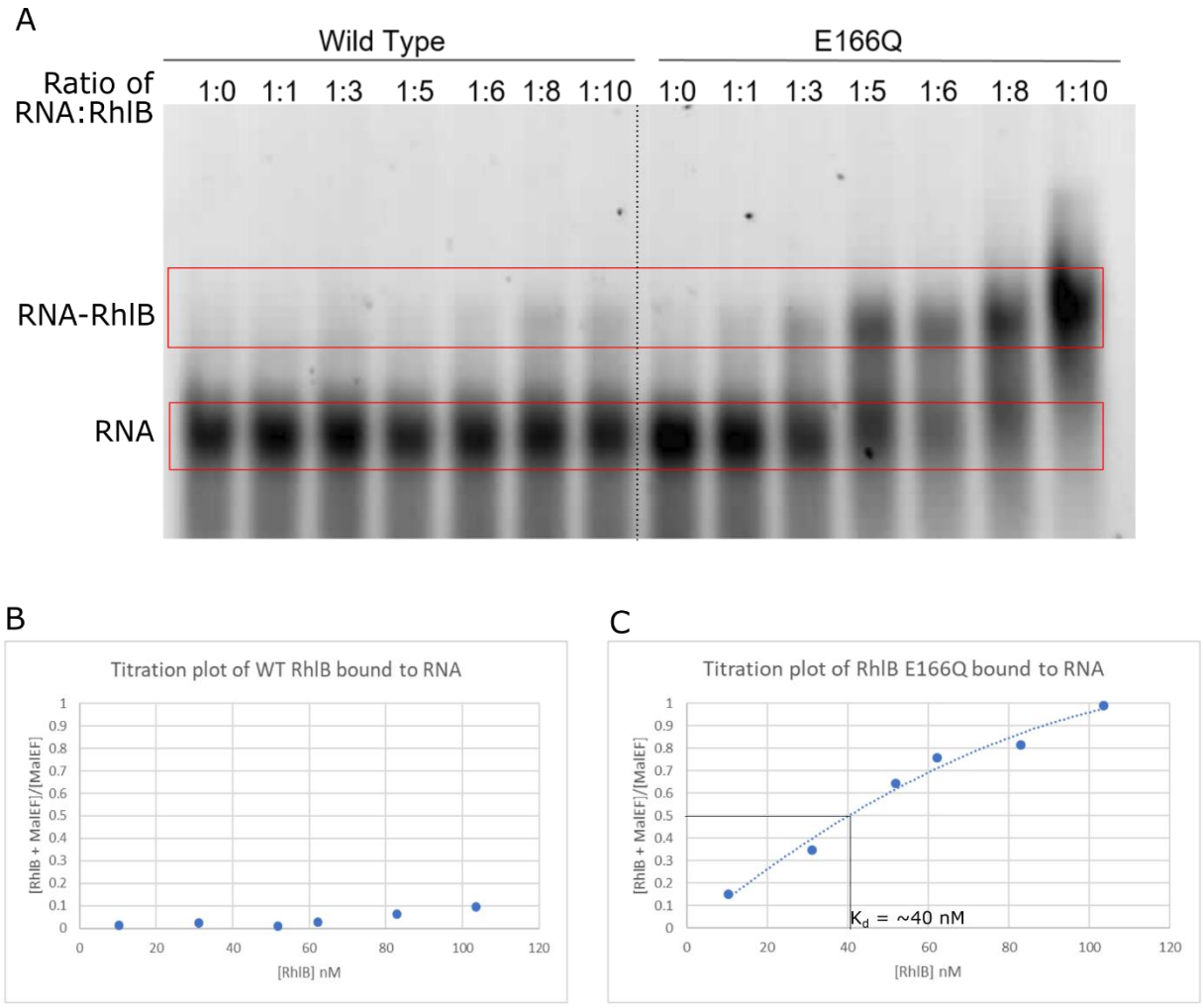


Figure 23: The effect of the E166Q mutation on RNA binding. A) An EMSA of the MalEF mRNA titrated with increasing amounts of RhlB visualized with SYBR-gold staining. The left 7 lanes are MalEF RNA titrated with WT RhlB, while the right 7 lanes are titrated with RhlB E166Q. Notable bands are highlighted in red. RhlB E166Q shows significantly higher levels of binding at higher concentrations as compared to the WT, which only shows a slight band of RNA-RhlB in the two rightmost lanes. Graphs B and C were plotted using data obtained from the image analysis software ImageJ. All datapoints were plotted as a ratio of bound to total RNA versus concentration of RhlB. The concentration of RNA used in this experiment was 10.36 nM. Graph B shows the titration plot of the WT RhlB binding to RNA. Graph C shows the titration plot of RhlB E166Q binding to RNA. When 50% of MalEF was bound to RhlB E166Q, the concentration of RhlB E166Q was around 40 nM. Thus, the K_d of the RhlB E166Q-MalEF complex can be estimated to be around 40 nM. Graph B displays an abnormally shaped titration curve, likely due to the low concentration of bound RNA. From the datapoints gathered, an accurate K_d cannot be estimated. However, note that the ratio of MalEF bound to WT RhlB remains below 10% at the RhlB concentration of 103.6 nM. Thus, it is likely that the Kd is even greater than 103.6 nM which would indicate a large discrepancy in binding affinity between the mutant and WT RhlB.

Using ImageJ, the EMSA gel was analysed in order to estimate the dissociation constant. A titration plot was used to determine the K_d, which is shown in Figure 23B and 23C. Figure 23B shows the titration plot used to calculate the K_d of WT RhlB and MalEF, while Figure 23C shows the plot used to calculate the K_d of RhlB E166Q and MalEF. The K_d calculated for the binding between RhlB E166Q and MalEF was around 40 nM. Though the K_d could not be calculated for binding between RhlB E166Q and MalEF, it is estimated to be above 103.6 nM. The mutation, therefore, increases RhlB's affinity to MalEF by at least 2 times. One characteristic of note is the irregular shape plotted in Figure 23B. Many points on the plot deviate from the expected shape of a titration curve, which may be due to the weak signal of the upper band in the EMSA gel. Thus, the calculated K_d should be interpreted with care. However, the low affinity of WT RhlB compared to RhlB E166Q is evident from the gel. Though the calculation of the precise increase in affinity is less reliable, there is little doubt that RhlB E166Q binds to RNA in a much tighter fashion than WT RhlB.

3.2.3 UV crosslinking and pulldown

Following the positive results obtained in the EMSA, our collaborators in Toulouse created a chromosomal point mutation in RhlB along with a 6xHis-FLAG tag at its N-terminus to enable in vivo studies of the 'RNA clamp'. Cells containing the mutation and tags were irradiated with 120 mJ/cm2 of 250 nm UV radiation. RhlB was first purified from other cell components with a denaturing nickel pulldown. A schematic of the experimental procedures is shown in Figure 24A. The pulldown following crosslinking of RhlB was successful, as the smear at the top of the Coomassie blue stained SDS-PAGE gel in lane 2 (Figure 24Biii) disappears following alkaline hydrolysis. Additionally, the RhlB band appears more prominent

in lane 3 following alkaline treatment, suggesting that RNA crosslinked to RhlB has been degraded by hydrolysis. Though crosslinking seems to occur, the pulldown yield was very low and not sufficiently pure for further analyses (Figure 24B). Attempts to further purify the crosslinked protein-RNA complex resulted in a complete loss of product. Thus, more experiments are needed to increase the yield of crosslinked RNA to allow for further analyses.

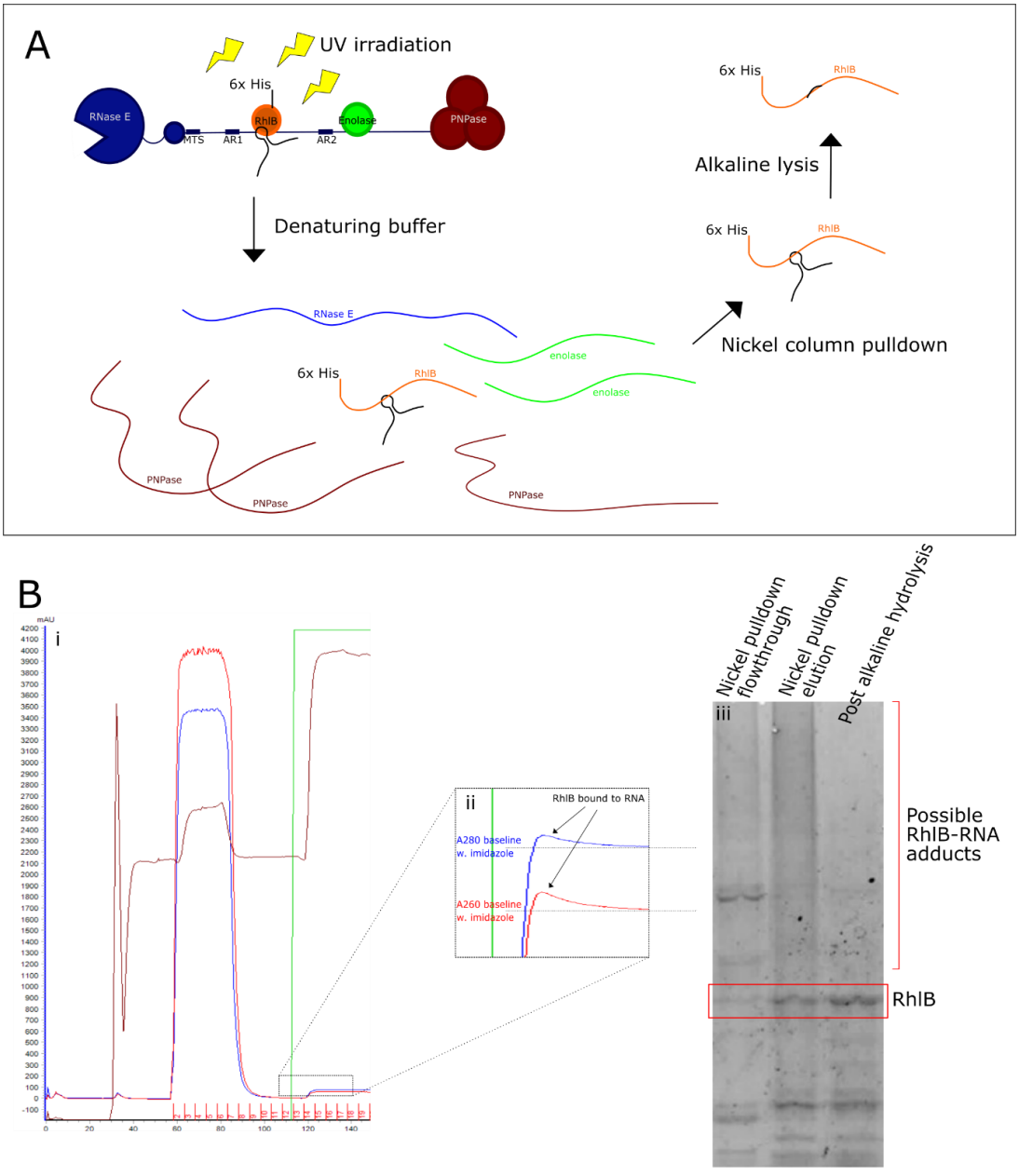


Figure 24: UV crosslinking and pulldown of RhlB. A) Schematic of pulldown experiment. A 6xHis tag was added at the N-terminus of RhlB. The cells were irradiated with UV light to induce covalent crosslinking between proteins and RNA. The cells were then lysed in denaturing buffer and passed through a nickel column to pulldown RhlB bound to RNA. The pulled down RhlB was then treated with NaOH to encourage RNA autolysis. RNA fragments covalently bound to RhlB should remain intact for further analysis. B) Chromatogram and SDS-PAGE gel of nickel pulldown. i) The chromatogram showing UV absorption during pulldown. The blue trace is A_{280} and the red trace is A_{260}. The green trace is the percentage elution buffer (see materials and methods) while the brown trace is conductance. Most of the RNA in the cell did not bind to the column, as the A_{260} is much higher than the 280 nm signal in the flowthrough. Though the amount is very small, some RhlB bound and was eluted (ii). The eluted RhlB can better be seen in panel (iii). Note that in lane 2, there is a smear on top of the gel which is not present in lane 3. Also note that the RhlB band in lane 3 appears slightly more prominent, suggesting that the smear is RhlB crosslinked to RNA.

3.3 Discussion

E. coli RhlB, when mutated in the DEAD box to DQAD, shows a similar clamping behaviour exhibited by mutated insect Vasa helicase. RhlB E166Q displays much higher affinity to MalEF than WT RhlB. The structural mechanism behind this change in affinity is likely to be the same as the Vasa helicase, as the DEAD box family have highly conserved core structures and motifs. RhlB E166Q is therefore also likely to exhibit the same capacity to form trapped complexes with target RNA *in vivo* shown by Vasa. With the UV crosslinking experiments, I have shown that RhlB can be covalently linked to RNA. Whether any transiently forming complexes are captured remains to be seen. Interestingly, the UV crosslinking experiments have shown a surprising characteristic of RhlB. A fraction of the total cellular RhlB is crosslinked to RNA, but there appears to be no visible RNase E band as well in the Coomassie stained gel (Figure 24B). As RhlB is bound to RNase E in close proximity to the RNA binding sites on RNase E, one could imagine that during the UV irradiation, RNA which is bound to both RhlB and RNase E would crosslink the two proteins together. It is possible that the probability of crosslinking RhlB to RNase E through an RNA is low, which results in crosslinked RNase E not being visualized by Coomassie staining. However, as RhlB theoretically should bind RNase E in a one to one ratio, and RNase E is much larger than RhlB, even a small percentage of RNase E crosslinked to RhlB was expected to be seen.

One possibility for the low number of RNase E crosslinks to RhlB seen in the results may be that the majority of RhlB in the cell is not attached to the RNA degradosome. The composition of the degradosome is highly variable, especially during stress conditions. It is known that in *E. coli* during cold stress RhlB is replaced with CsdA (Prud'homme-Géńreux et

al, 2004). The displaced RhlB may, therefore, simply become a part of the larger cellular pool of RhlB. There have been no experimental data supporting a potential cytosolic function of RhlB if it is displaced from the degradosome. Though the possibility exists that RhlB may function independently from the degradosome, this scenario seems unlikely given that RhlB and RNase E co-localize within the cell, and very few RhlB molecules are found away from the membrane (Moffitt et al, 2016). Thus, it is likely that the missing RNase E crosslinking may have another cause.

Another possibility is the involvement of sRNA binding by RhlB. It may be that the primary RNA binding partner of RhlB is not large polycistronic mRNAs, but rather small regulatory sRNAs. This hypothesis is supported by the fact that Hfq is known to bind the RNA degradosome at the site of RhlB binding (Ikeda et al, 2011). Thus, the possibility arises that Hfq bound sRNAs can be recruited to the degradosome through RhlB binding. RhlB, when bound to RNase E, turns over ATP quite rapidly (Chandran et al, 2007). This characteristic could potentially "mask" RhlB's involvement with sRNAs, since their interactions would be limited to one ATPase cycle. With the RhlB E166Q mutation forming the RNA clamp, sRNA association with RhlB could be shown to be a larger percentage of RhlB's RNA binding partners. Though more experiments are necessary to test these hypotheses, RhlB's involvement in RNA degradation has been shown to be more complex than previously imagined. Further investigations into RhlB function will be integral to a deeper understanding of the intricacies of RNA degradation.

3.4 Materials and methods

3.4.1 Cloning

Site-directed mutagenesis was performed on RNase Helicase B (RhlB) to introduce a mutation at the DEAD box site (aa165-169). Specifically, the glutamic acid at residue 166 was to be mutated into a glutamine (E166Q), requiring the mutation of codon GAA at nucleotides 496-498 into CAG. A pair of primers were designed to cover enough of the mutation site to be specific as well as introduce the point mutation (Figure 25C). Primers were made by Sigma. Primer pellets were resuspended and diluted to the concentration of 100 μM and stored in -20 °C. PCR was performed with a Phire II kit (Thermo Scientific) according to the provided protocol: PCR mix was made with 1x Phire reaction buffer, 10 ng template DNA, 200 μM dNTPs, 0.5 M forward and reverse primers, and 1 μL Phire hot start polymerase. The template DNA used was *rne* 603-850 and *rhlB* in the pRSFDuet-1 vector, which was kindly provided by Dr. Heather Bruce. The PCR temperature cycle was programmed as follows: Initial denaturation at 98 °C for 2 min, followed by 32 cycles of denaturation at 98 °C for 5 s, annealing at 60 °C for 5 s, extending at 72 °C for 1 min, and then a final extension step at 72 °C for 10 min. Completed PCR reactions were purified with a Qiagen PCR purification kit. Purified plasmids were transformed into competent *E. coli* DH5α cells and plated on kanamycin plates. Plated cells were incubated at 37 °C overnight (16 h). Multiple colonies were picked and grown individually in 5 mL of Luria-Bertani (LB) media with 30 μg/mL of kanamycin overnight at 37 °C. Overnight cultures were harvested by centrifugation at 20,000 xg for 15 min. Plasmids were isolated with Thermo GeneJET Plasmid Miniprep Kit. Isolated plasmids were sequenced by Sanger sequencing at the Department of Biochemistry in Cambridge to check for the successful

site-directed mutation at the DEAD box (Figure 25D). Successfully mutagenized plasmids were stored at -20 °C.

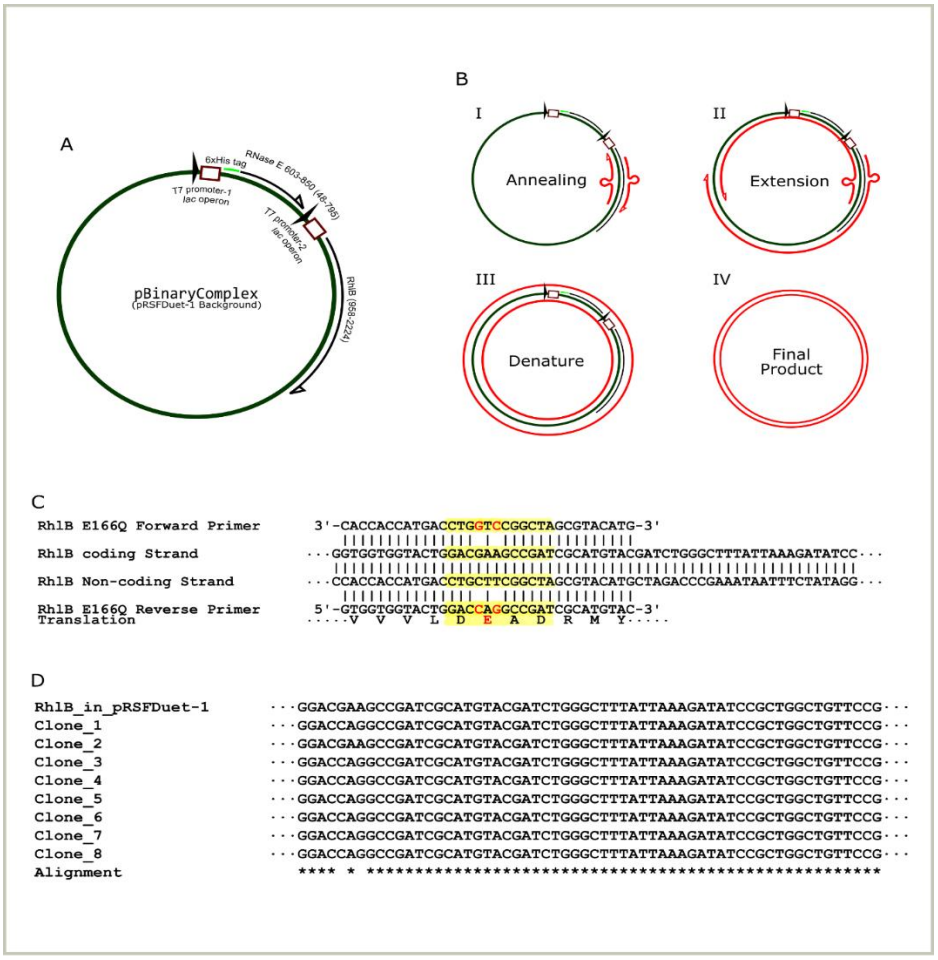


Figure 25: An overview of the cloning process and its associated sequences. A) A cartoon schematic of the initial plasmid, pBinaryComplex, before the mutagenesis process. This plasmid was made with the background of pRSFDuet, which also confers kanamycin resistance to the transformed bacteria. B) An overview of the site-directed mutagenesis process. The new primers and synthesized DNA is in red, while the template DNA is coloured green. C) The primers designed for this process. The DEAD box is highlighted, while the introdced mutations were coloured red. D) The final sequence obtained via Sanger sequencing. The top sequence is the template sequence which was used in the site directed mutagenesis, while the other 8 were mutagenized samples. The last row is the alignment, where a * represented perfect alignment and no * denotes the absence of alignment.

3.4.2 Expression and Purification of ternary complex (RhlB + RNase E 603-850 + enolase)

Competent BL21 cells were co-transformed with pBinaryComplex (Figure 25A) and an a plasmid that encodes enolase kindly provided by Dr. Heather Bruce at the same time. Transformed cells were plated on kanamycin and carbenicillin-containing agar plates and incubated at 37 °C overnight (16 h). Starter culture was made by inoculating 20 mL of LB media with 30 μg/mL of kanamycin and carbenicillin with one colony of transformed cells. Inoculated media was grown overnight (16 h) at 37 °C with aeration. 2 x 1 L of LB 1 media with 30 μg/mL of kanamycin and carbenicillin was inoculated with 2 x 10 mL of starter culture. Inoculated 1 L cultures were grown at 37 °C until cells reached mid-log phase (OD_{600} = 0.6). 1 mM IPTG was added to 1 L cultures to induce protein expression. Cultures were harvested 5 h after expression induction via centrifugation at 4,000 xg for 20 min at 5 °C using a Beckman centrifuge with a 4.2 rotor. Cell pellets were stored at -80 °C prior to lysis. On the day of purification, frozen cells were thawed in ternary complex lysis buffer (for every gram of the dry cell mass, three to five times the mass of buffer is added). Thawed cells were lysed via high pressure homogenization using an Avestin Emulsiflex C5. Thawed cells were passed through the homogenizer at 5000 psi four to six times, or until DNA was sheared. Cell debris was separated from soluble proteins by centrifugation at 30,000 xg for 30 min. Meanwhile, a nickel column was equilibrated in ddH_2O (4-5 column volumes) and ternary complex nickel binding buffer (Table 4, 4-5 column volumes). Clarified cell lysate was then loaded onto the equilibrated nickel column and the column was then washed with 4 column volumes (CV) of ternary complex nickel binding buffer (Table 4), or until 280 nm reached baseline. Bound proteins were eluted with 20 mL of ternary complex nickel elution buffer, collected in 2 mL fractions. All

peak fractions were visualized on an SDS-PAGE gel. Relevant fractions were pooled and dialyzed in dialysis buffer (Table 4) for 4 h at 4 °C. Dialyzed proteins were loaded onto an equilibrated heparin column. The loaded column was washed with 4 x CV of ternary complex heparin binding buffer (Table 4) or until the 280 nm absorbance signal reached a stabilized baseline value. Bound proteins were eluted from the heparin column with a gradient from 0-1 M NaCl over 120 mL. Eluate was collected in 2 mL fractions. Peak factions were visualized on an SDS-PAGE gel. Fractions containing clean ternary complex were concentrated and loaded onto a S200 column equilibrated with ternary complex S200 buffer (Table 4). Fractions of 0.3 mL were collected. Peak fractions were analyzed with SDS-PAGE and pure ternary complex was stored in -80 °C.

Table 4: List of buffers used for ternary complex purification.

Buffer Name	**Buffer composition**
Ternary complex lysis buffer	Tris pH 7.5 (50 mM), NaCl (250 mM), KCl (100 mM), $MgCl_2$ (10 mM) Imidazole (2 mM), β-mercaptoethanol (50 µM), 1x protease inhibitor tablet per 50 mL
Ternary complex nickel binding buffer	Tris pH 7.5 (50 mM), NaCl (250 mM), KCl (100 mM), $MgCl_2$ (10 mM) Imidazole (2 mM), β-mercaptoethanol (50 µM)
Ternary complex nickel elution buffer	Tris pH 7.5 (50 mM), NaCl (250 mM), KCl (100 mM), $MgCl_2$ (10 mM), Imidazole (300 mM), β-mercaptoethanol (50 µM)
Dialysis buffer	Tris pH 7.5 (50 mM), NaCl (250 mM), KCl (20 mM), β-mercaptoethanol (50 µM)
Ternary complex heparin binding buffer	Tris pH 7.5 (50 mM), NaCl (250mM), $MgCl_2$ (10 mM) β-mercaptoethanol (50 µM)

Ternary complex heparin elution buffer	Tris pH 7.5 (50 mM), NaCl (1 M), KCl (250 mM), $MgCl_2$ (10 mM), β-mercaptoethanol (50 μM)
Ternary complex S200 buffer	Tris pH 7.5 (50 mM), NaCl (200 mM), KCl (250 mM), $MgCl_2$ (10 mM) β-mercaptoethanol (50 μM), glycerol (10 % v/v)

3.4.3 MalEF *in vitro* transcription

E. coli K12 total DNA was kindly provided by Dr. Kasia Bandyra. Two sets of PCR primers were designed, one to amplify the region of interest to provide a template for the second set, used to make a DNA template for in-vitro transcription with T7 polymerase. PCR mix was first made with 1 x Phire reaction buffer, 10 ng template DNA, 200 μM dNTPs, 0.5 μM of both malE_malF_FL_Nested_FW and malE_malF_FL_Nested_RV (Table 5), and 1 μL Phire hot start polymerase. The template DNA used was *E. coli* K12 genomic DNA. PCR temperature cycle was programmed as follows: Initial denaturation at 98 °C for 2 min, followed by 32 cycles of denaturation at 98 °C for 5 s, annealing at 60 °C for 5 s, extending at 72 °C for 1 min, and then a final extension step at 72 °C for 10 min. Amplified *malEF* PCR template DNA was purified with a Qiagen PCR purification kit. The PCR reaction was repeated with the previous product being the template, using malEmalF_IVT_FW and malEmalF_IVT_RV as primers (Table 5). PCR mix was made with 1 x Phire reaction buffer, 10 ng template DNA, 200 μM dNTPs, 0.5 μM forward and reverse primers, and 1 μL Phire hot start polymerase. The template DNA used was purified in the previous PCR reaction. PCR temperature cycle was programmed as follows: Initial denaturation at 98 °C for 2 min, followed by 32 cycles of denaturation at 98 °C for 5 s, annealing at 60 °C for 5 s, extending at 72 °C for 1 min, and then a final extension step at 72 °C for 10 min. Amplified *malEF* IVT template DNA was purified with a Qiagen PCR

purification kit. Purified IVT template DNA was used to transcribe MalEF RNA. *In vitro* transcription (IVT) reaction mix was made with 1 x reaction buffer (50 mM Tris pH 7.5, 10 mM NaCl, 5 mM $MgCl_2$, 200 μM spermidine), 100 μM rNTPs, 3 μg of T7 polymerase, 1 μg of template DNA in 4 mL of RNase free H_2O. IVT reaction mix was incubated at 37 °C for 4 h. RNase Free DNase I was then added to the reaction mix and incubated for 30 min at 37 °C. All reactions were stopped with the addition of RNase free 10 mM EDTA. Stopped reactions were run on a 4 % urea sequencing gel for 5 h. The gel was illuminated with short wavelength UV light to identify the relevant RNA band, which was then excised. RNA was electroeluted from the gel and precipitated with 300 mM sodium acetate pH 5.3 and 3 volumes of ethanol in -20 °C overnight (16 h). Precipitated RNA was harvested by centrifugation. The RNA pellet was dried and stored in -20 °C.

Table 5: List of primers used in MalEF IVT

Primer name	Primer sequence
malE_malF_FL_Nested_FW	CTGGCGGAGTGTCATTCATCCGTTTCTCAC
malE_malF_FL_Nested_RV	GCAGCGCAATCCCACCCAGATACGC
malEmalF_IVT_FW	CATTAGTAATACGACTCACTATAGGGATGAAAATAAAAACAGGTGCACG CATCCTCGCATTATC
malEmalF_IVT_RV	TTAATCAAACTTCATTCGCGTGGCTTTCAGGTTCACT

3.4.4 Titration of MalEF by ternary complex (RhlB + RNase E 603-850 + enolase)

Increasing concentrations of ternary complex were added to MalEF at room temperature. Each titration condition was loaded in one lane of the 1 % agarose gel. The gel was run in 1 x TBE at a fixed voltage of 100 V at 4 °C for 2 h. The gel was visualized by UV illumination after staining with 5 μL of SYBR gold in 100 ml gel running buffer for 10 min under a foil cover. Relative concentrations of protein-RNA complexes were calculated using ImageJ.

3.4.5 UV irradiation and pulldown

Cell strains containing the RhlB E166Q mutation as well as an N-terminal 6 x His + FLAG tag on RhlB were obtained from our collaborators in Toulouse and stored at -80 °C as a glycerol stock. The glycerol stock was swabbed with a cell picked and streaked on LB agar plates with 30 μg/mL kanamycin and grown at 37 °C over 16 h. One colony was picked and grown in liquid LB media with 30 μg/mL kanamycin at 37 °C over 16 h with aeration. 2 L of LB media with kanamycin was inoculated with 5 mL of starter culture and incubated at 37 °C with aeration. When OD_{600} reached 0.6, cultures were harvested by centrifugation at 4000 rpm in a Beckman centrifuge with a 4.2 rotor for 20 min at 5 °C. Cell pellets were resuspended in 30 mL irradiation buffer (Table 6). Resuspended cells were frozen in liquid nitrogen, forming 60 μL frozen cell drops. Frozen cell drops were spread out in one layer and irradiated with 120 mJ/cm^2 of UVC radiation (200 nm to 280 nm). Irradiated cells were thawed and harvested by centrifugation at 4000 rpm for 20 min. Cell pellets were resuspended in denaturing nickel

column binding buffer (Table 6). Resuspended cell pellets were lysed using the freeze-thaw method. The lysate was clarified by centrifugation at 30,000 x g for 30 min. Clarified lysate was loaded onto a nickel column. Lysate was washed with 4 x CV of denaturing nickel column binding buffer or until the UV trace returned to its baseline value. Bound proteins were eluted with 4 x CV of denaturing nickel column elution buffer (Table 6). Eluted proteins were collected in 0.3 mL of fractions. Alkaline hydrolysis was performed by addition of 0.5 µL of 5 M NaOH to 20 µL of protein and incubating at 25 °C for 10 min. Relevant fractions and alkaline lysis products were analyzed by SDS-PAGE.

Table 6. List of buffers used for UV irradiation and pulldown

Buffer Name	Buffer composition
Irradiation buffer	Tris pH 7.5 (50 mM), NaCl (250 mM), KCl (100 mM), $MgCl_2$ (10 mM), β-mercaptoethanol (50 µM)
Denaturing nickel column binding buffer	Tris pH 7.5 (50 mM), NaCl (250 mM), KCl (100 mM), $MgCl_2$ (10 mM) Imidazole (2 mM), urea (6 M), β-mercaptoethanol (50 µM)
Denaturing nickel column elution buffer	Tris pH 7.5 (50 mM), NaCl (250 mM), KCl (100 mM), $MgCl_2$ (10 mM) Imidazole (300 mM), urea (6 M), β-mercaptoethanol (50 µM)

4 STRUCTURAL STUDIES OF HFQ BOUND TO CHIX

4.1 Introduction

Following biophysical studies of RhlB E166Q binding to RNA, I explored the use of the point mutation to create a stable complex for structural analyses. As shown in the previous chapter, RhlB E166Q is much more binds RNA tighter than the wild type enzyme and therefore could create a more robust ternary complex suitable for structural characterization. Although the N-terminal portion of RNase E is well studied, the C-terminal of RNase E and the overall organization of the degradosome including its complex with RhlB has eluded structural determination. Thus, using the RNA clamping RhlB E166Q variant, I hoped to create a more stable complex for structural studies. As the degradosome's C-terminal portion is highly flexible and adopts a variety of conformations which may present difficulties for crystallisation, I opted to use the newly popularized technique of cryo-electron microscopy (cryoEM) for structural analyses of RhlB and its subassemblies with RNase E.

In conventional microscopy, the wavelength of light imposes a hard limit on the resolution limit of the imaged specimen. Objects smaller than the wavelength cannot be imaged due to diffraction limitations. For visible light, this means that any feature smaller than 200 nm in length cannot be resolved (Kruger et al, 2000). Super-resolution methods using fluorophores (such as STED, PALM, STORM and other approaches) can overcome the diffraction limitation but are still limited to a 20 nm resolution (Betzig et al, 2006; Hell & Wichmann, 1994; Rust et al, 2006). Structural studies of biological macromolecules require the resolution to be in the

order of 1 Å, which is 2000 x smaller than the resolution limit of light microscopes. To solve this problem, electrons are used for visualization (Kruger et al, 2000). At the energy of 200 keV, an electron has a corresponding wavelength of 2.74 pm, which is more than 30 x smaller than the resolution required for determining the atomic structure of a protein (Spence). Since electron microscopes are so powerful and allow for imaging of atoms, they have been in use for decades in the material sciences. The electron microscope's adaptation for use in the biological sciences has been deterred largely by the fragility of biological samples (Kruger et al, 2000). With such powerful energy beams as the illumination source, biological samples were easily damaged beyond recognition. However, with the invention of the new sample preparation techniques available with cryo-EM, biological samples may now be imaged using the electron microscope by minimising the damage caused by the electron beam (Knapek & Dubochet, 1980). Cryo-EM has been the primary method by which the images in this chapter have been obtained. A more detailed explanation surrounding the hardware, theory, and sample preparation is now presented.

The principles of image formation in electron microscopy are very similar to those of light microscopy (Spence). Thus, the design of electron and light microscopes also share common features (Figure 26). Electrons travelling through the microscope must travel through vacuum to prevent undesired scattering by gas particles prior to reaching the detector, and the entire chassis is therefore maintained under low pressure (Spence). At the top of the microscope is an electron source. Broadly speaking, there are two different types of sources which all electron microscopes use to generate an electron beam, thermionic emission sources and field emission guns (FEG) (Houdellier et al, 2015; Iqbal & Fazal-e-Aleem, 2005). Electrons generated from thermionic sources are pulled from a v-shaped tungsten wire which is heated to

lower the energetic barriers needed for electrons to escape (Iqbal & Fazal-e-Aleem, 2005). In a field emission gun, no heat is applied to the electron source. Instead, electrons are directly pulled from a sharp crystalline point on the gun with a strong electric field (Houdellier et al, 2015). The difference in the electron beam produced by thermionic sources versus field emission guns is analogous to the light source provided by a light bulb and a laser respectively, as FEGs produce both more powerful and coherent electron beams (Houdellier et al, 2015). Thus, it is no surprise that currently, FEGs are the more popular choice for cryo-EM.

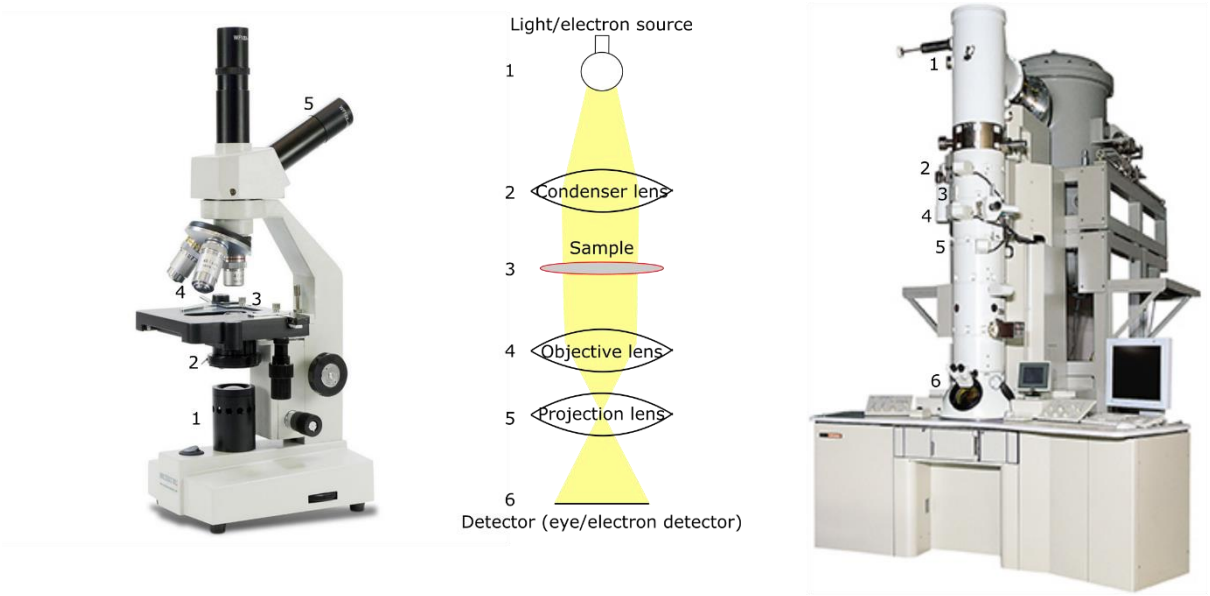


Figure 26: Architecture of a generic transmission electron microscope (image copied from Torrisi, 2011) vs a light microscope (HOME-DUAL__40784.1549918163, 2019). In both, the electron/light source creates the beam, which passes through the condenser lens and illuminates the sample, where either photons or electons are scattered. Then, the beam now carrying an image of the sample passes through the objective and projection lens and is detected by either the eye (light microscope) or an electron detector (electron microscope).

Following the electrons beam's creation, multiple magnets are used as lenses to focus the beam onto the sample. The electrons are scattered by the sample and focused by the objective lens. After focusing, the electrons' positions must be detected. Historically, electron detection has been done by photographic film and charged-coupled device (CCD) detectors (Bai et al, 2013). These techniques require the conversion of electrons into photons before detection, thus lowering the signal to noise ratio. In 2009, a novel detector was reported that could directly detect electrons without the requirement for photoconversion, and the data collected with this device showed significantly increased signal to noise ratios (McMullan et al, 2009). Since then, a variety of direct electron detectors have been developed and put into use for single particle analysis to achieve high resolutions with lower sample damage by permitting the use of low electron dose on the specimen (McMullan et al, 2016).

Electron beams are highly energetic and can deposit some of their energy into specimens in a destructive way. As the name suggests, cryo-EM protects biological samples from electron beam damage by using cryogenic conditions. As crystalline ice scatters electrons strongly, the specimen must be frozen in a thin layer of vitreous ice to allow for adequate electron scattering by the sample. The sample of interest is applied onto a cryo-EM grid, made from a metal, usually copper or gold, with or without a carbon support. Once the sample is deposited, the grid is blotted to remove any excess liquid. Thus, the only liquid remaining is suspended in the grid by water's adhesive forces. After blotting, the grid is rapidly plunge frozen in liquid ethane, forming a thin layer of vitreous ice with the protein of interest in suspension and hopefully in random orientations. Once the sample is prepared, it must be stored in liquid nitrogen until imaging (Figure 27).

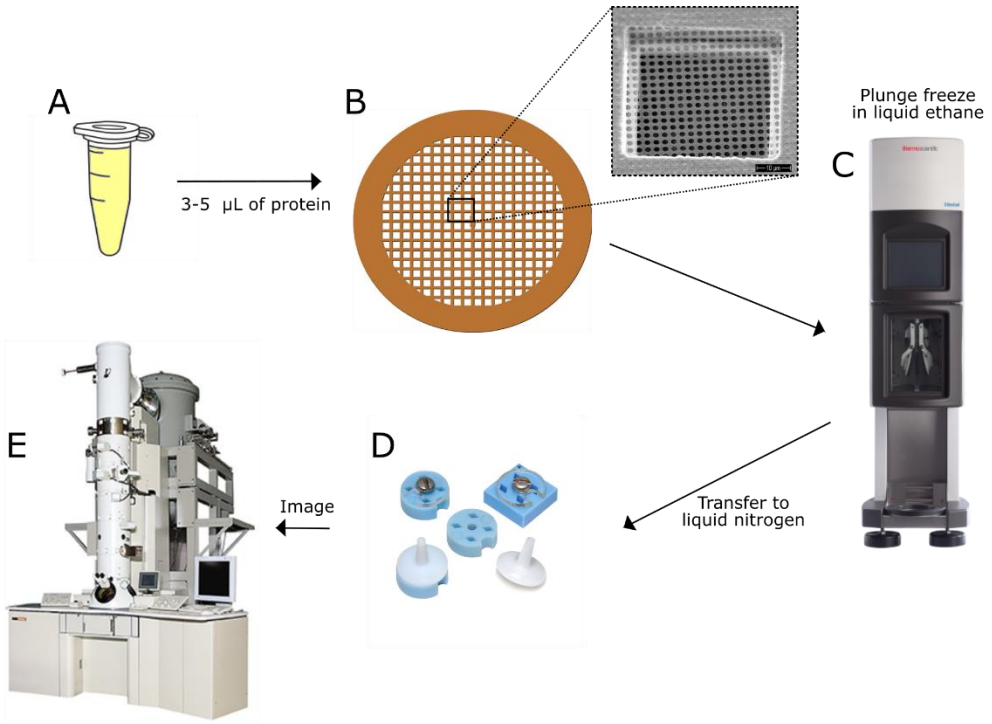


Figure 27: Cryo-EM sample preparation flow chart. 3-5 µL of the sample of interest (is deposited on a grid (B) (copper-tem-grid-500x500, 2019; Quantifoil_R12_13-5, 2019), blotted, then plunged into liquid ethane. Most commonly, blotting and plunge freezing is performed by the vitrobot (C) (Vitrobot_500x500, 2019) which controls the temperature and humidity a grid experiences as well as the blot time and blot force. The grid is then transfered from liquid ethane to liquid nitrogen and into grid boxes (D) (g3727_cryo-grid-boxes-v2_1200px, 2019). Prepared grids are stored in liquid nitrogen before imaging in the microscope (E) (Torrisi, 2011).

In theory, electron microscopes are powerful enough to resolve single particles in one image. In practice, the images obtained from electron microscopes are very noisy because the electron dose is kept to a minimum to visualise the particles without destroying them. To optimise the contrast and thus allow for the visualization of the particles of interest, the images are collected slightly out of focus, so that the image obtained is analogous to that formed by the phase contrast method used in optical microscopes. The images of the individual particles are

difficult to interpret without processing, which includes corrections for the out of focus effects. To compensate for the low signal to noise ratio, many images of the same particle must be obtained to reconstruct a 3D structure in high resolution. The images collected by the cryo-EM are 2D projections of a 3D object. The 3D reconstruction operations are performed in Fourier space (Figure 29C) to enable frequency filtering and application of the contrast transfer function, hereby referred to as CTF, to correct for defocus and optical aberrations. Briefly, Fourier space is an alternate way of describing the values in an image. Instead of mapping the square of the amplitude of an electron wave (which is measured by the direct electron detectors), Fourier space maps the frequency at which each wave occurs in a 2D image. The advantage of Fourier space is that different functions can be added or removed from the image through convolution functions to process and refine the image. For example, low-pass filtering can be done easily in Fourier space simply by blocking out the information corresponding to the higher frequency waves, which are the furthest away from the origin in Fourier space. The Fourier transform of a 2D projection of the object corresponds to a 2D slice of the Fourier transformed 3D object. Thus, by having a complete dataset of 2D projections from different angles, a 3D reconstruction of the object can be calculated in Fourier space. The reconstructed object can then be back-transformed into real space as the object of interest.

Images collected by the detector of an electron microscope are imperfect projections of the "perfect" sample, being distorted based on the defocus applied and the electron's interference fringes. Each point on a defocused image is slightly more spread out than the ideal image of that point. Thus, to calculate the ideal image, the spread caused by defocus must be removed. The spreading shown in the collected images can be described by the point spread function. In real space, the point spread function is difficult to remove and requires large

amounts of computational power. However, in Fourier space, removal of the point spread function is much simpler. The Fourier transform of the point spread function is the CTF, and CTF correction is processed in a much faster way in Fourier space. Calculating and correcting for the CTF is important for obtaining images of the highest possible quality by electron microscopy (Figure 28A).

To collect a complete dataset without missing information, samples must be imaged at different defocus levels. As defocus distorts the "perfect" projection of the sample by spreading the image, it also changes the interference in the electron waves which eliminates information. The CTF illustrates the amount of information available as well as the information which has been eliminated as a function of spatial frequency (Figure 28B). At certain spatial frequencies, no information about the sample is transferred to the collected image. In the CTF, these frequencies are shown as the points where the amplitude crosses the x-intercept. Without the information in these spatial frequencies, the data collected is incomplete. This apparent defect is circumvented by collecting images using different defocuses, which changes these intercepts to different spatial frequencies (Figure 28C). By moving the intercepts, a complete dataset which contains information in all relevant spatial frequencies can be collected.

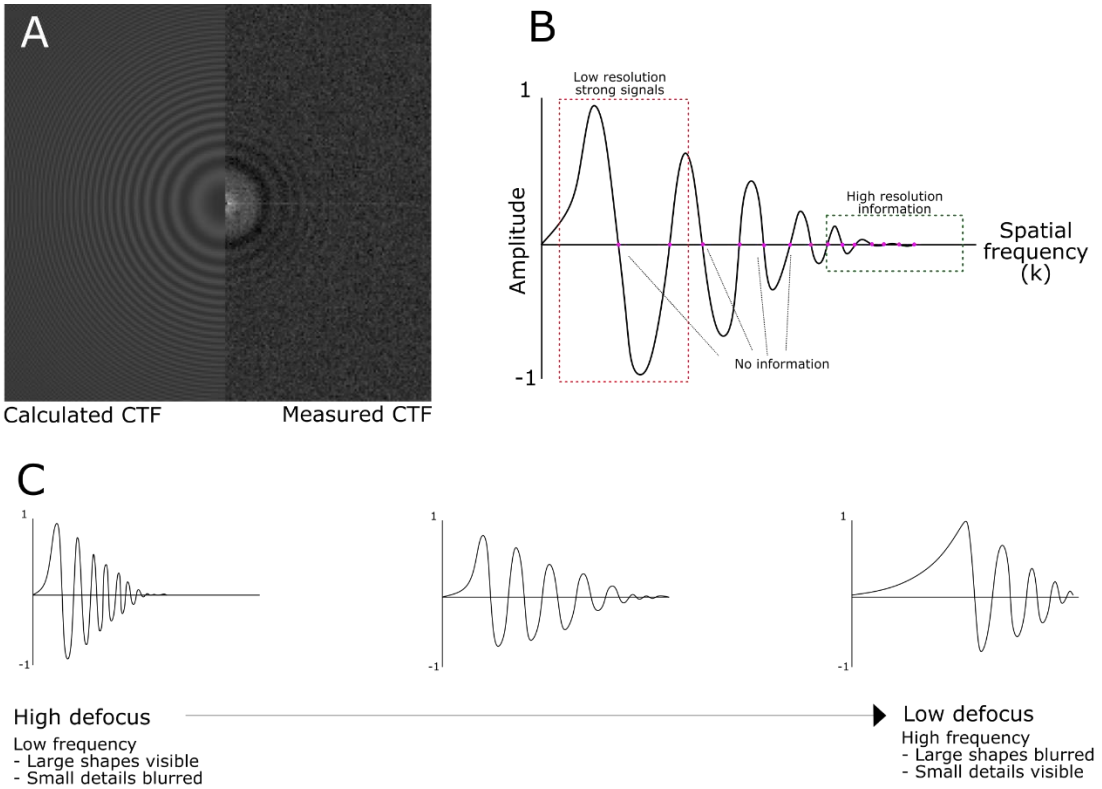


Figure 28: Contrast transfer functions. A) An image of a calculated CTF from a micrograph obtained from Cryo-EM. B) A 1D graph of a CTF. The x-axis denotes the spacial frequency while the y-axis denotes the amplitude, or the amount of signal contributed by each frequency. In a typical image, at low spatial frequencies (and therefore low resolution, as resolution is proportional to spatial frequency), amplitudes are high while at high spatial frequencies amplitudes are low. C) The effect of defocus of contrast and resolution. At high defocus, low frequencies have high amplitude, meaning that shapes will be more visible. However, small details will be blurred as there is very little high frequency information. On the other hand, images taken closer to focus contain more higher resolution information but less low resolution information. As a result, large shapes are difficult to disseminate. Thus, a wide range of defocuses are necessary for a complete reconstruction of a protein from cryo-EM images.

After the correction of aberrations, particles of interest must be aligned to boost the signal to noise ratio. One of the most popular software for aligning particles and 3D

reconstruction is Relion, which utilizes Bayesian statistics to accurately align multiple images and generate 2D and 3D models made from the images obtained (Fernandez-Leiro & Scheres, 2017). The Relion software creates a data processing pipeline and applies several algorithms for pre-processing – such as CTF correction – which are wrapped in the package along with Relion's own particle autopicking, 2D classification, 3D classification, 3D refinement and post-processing programs (Rohou & Grigorieff, 2015; Scheres, 2018; Zhang, 2016; Zheng et al, 2017). To obtain robust 2D projections of the model, particles are classified using a weighted maximum likelihood estimation (Figure 29B). Each image is aligned using a gaussian probability distribution to the orientation parameters of the 2D image. The image is then weighted against the probability obtained and used to boost the signal of the original image. This process is then performed iteratively until convergence to ensure that the model generated is the most likely structure from which the images are obtained (Scheres, 2012). This process is performed for each 2D projection slice, which then can be used for reconstructing a 3D map of the object. The *de novo* coulomb potential map generated may then be fit to the molecules of interest. One benefit of this technique is the potential to generate multiple maps from one set of images, meaning that multiple conformations of the molecule of interest captured in ice can be imaged at the same time. Thus, flexible regions and sample heterogeneity can be accounted for using cryo-EM, giving structures which are highly relevant to *in vivo* functions of individual proteins.

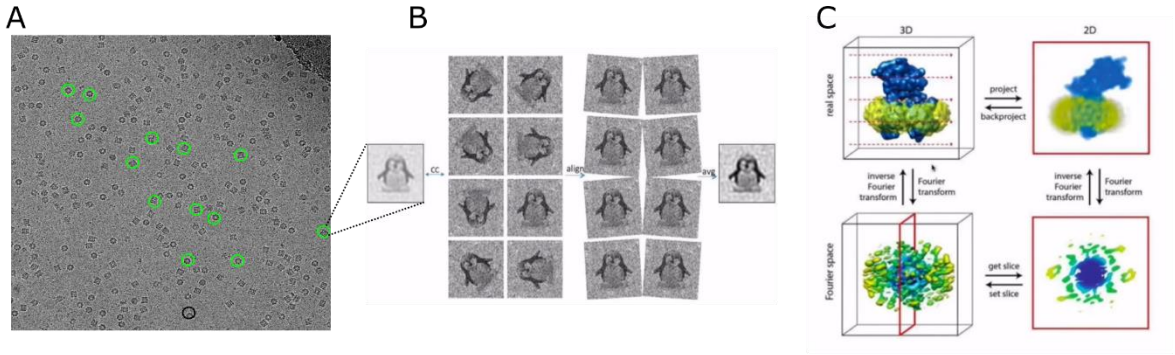


Figure 29: Image processing of cryo-EM. Images B and C obtained from Scheres, 2015. A) The micrograph obtained after imaging a grid. Images obtained from the microscope often need to be corrected for motion and spherical abberations which are due to the imperfections in the magnetic lenses and electron induced protein movement. Following various corrections, particles can be extracted from the micrographs shown as the green circles. Extracted particles can then be aligned to boost signal to noise ratio (B). Following alignent, 2D projections of the image can then be reconstructed into a 3D map of the protein of interest (C)

Recently, a SAXS (small angle X-ray solution scattering) envelope of a C-terminal segment of RNase E bound to RhlB and enolase was reported (Bruce et al, 2018) which indicates that the C-terminal of RNase E is more compacted than expected for a completely disordered region, suggesting partial folding of the C-terminal tail of RNase E. As cryo-EM has been shown to be useful in imaging flexible and disordered regions, I aimed to investigate the structure of the semi-compacted complex comprising the C-terminal region of RNase E (residues 603-850) bound to RhlB, enolase, Hfq and ChiX, hereby denoted as the supercomplex. Larger proteins and complexes align better in Relion, which was another consideration for choosing the supercomplex for this study. Utilizing the RhlB E166Q mutation described in the previous chapter was anticipated to help capture the ChiX RNA with the helicase clamped to the oligonucleotide. I hoped that, upon binding to RNA, the C-terminal domain's RNA binding sites would display more consistent structures which may simplify the structure determination process. In the introduction (chapter 1 section 4, page 36 - 37), ChiX

was discussed to be a class II sRNA, which binds both sides of Hfq. I envisioned that, in addition to understanding the organization and structure of the RNase E C-terminus, I might also gain insight into the novel association between a class II sRNA and Hfq.

4.2 Results

Hfq and the ternary complex (consisting of RhlB E166Q, enolase, RNase E 603-850) were purified. The purification of Hfq is illustrated in Figure 30, and the procedure for preparing the ternary complex is shown in Figure 22 from the previous chapter. The ternary complex was added in a 1:1:1 ratio to purified ChiX and Hfq to form a supercomplex that could be visualised by EMSA (Figure 32). The supercomplex was applied to a copper quantifoil grid then plunge frozen in liquid ethane using a vitrobot device (Figure 32). The resulting grid was imaged using a Titan Krios instrument fitted with a Falcon II detector.

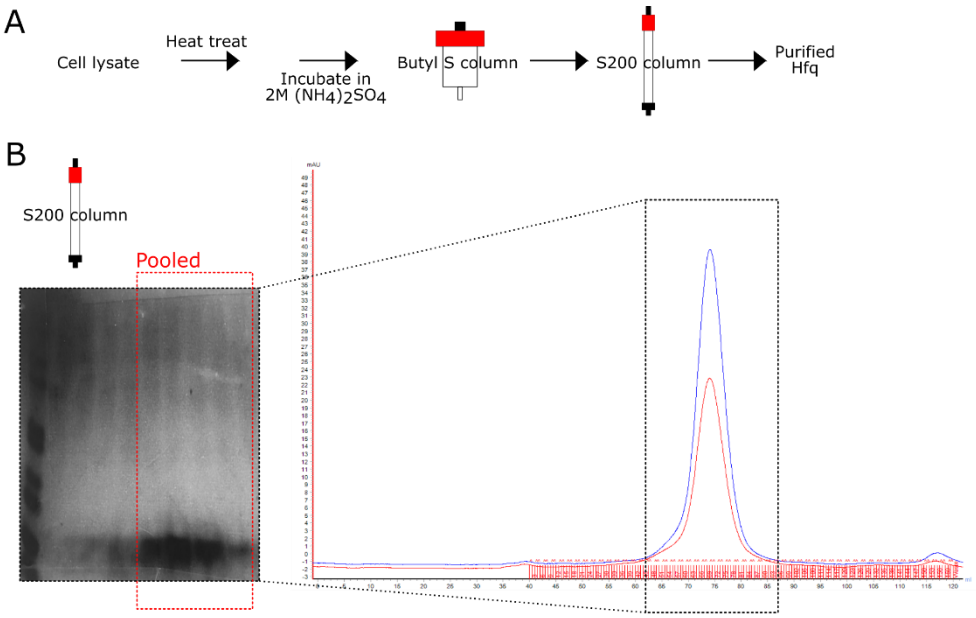


Figure 30: Purification of Hfq. A) Schematic of Hfq purification. After cell lysis, clarified cell lysate was first heat treated for 60 min. Heat treated lysate was then clarified by centrifugation. After heat treatment, an equal volume of 4 M ammonium sulfate wass added to the lysate and incuated for 60 min. Then the proteins were purified by hydrophobic interaction followed by size exclusion. B) UV absorbance trace of size exclusion column as well as its resulting SDS-PAGE gel. The blue line denotes UV 280 while the red line denotes UV 260. The scale for both profiles were identical. The ratio of 280 to 260 indicates that the sample was free from nucleic acid contaminants.

149 images were collected in total (Figure 31A). Particles were picked with Relion autopicking software and then manually selected via particle sorting in Relion. Cleaned particles were classified by 2D classification in Relion and poorly aligning particles were discarded. This process left about 80,000 particles for 3D reconstruction. During 2D classification, it was noted that the classified particles were too small to be supercomplex particles according to the SAXS dimensions published by Bruce *et al* (2018). Nonetheless, as the particles could be an interesting subcomplex, I decided to further analyze the classes obtained. Many subclasses of the supercomplex could in principle fit the model, but the most

likely candidate was Hfq bound to ChiX. As Hfq is a hexagonal ring, the donut shape seen in the 2D classes (Figure 31 C) suggested that Hfq was likely to be present. The 3D reconstruction was however found to be limited to poor resolution and too small to fit both Hfq and ChiX. As enolase was also previously observed to fall off the ternary complex and adhere to the air-water interface during grid preparation, the possibility of this smaller subdomain being an enolase dimer was also considered. The obtained resolution of the envelope was too low to identify the exact subcomplex, therefore more samples were prepared and imaged using different sample preparation techniques.

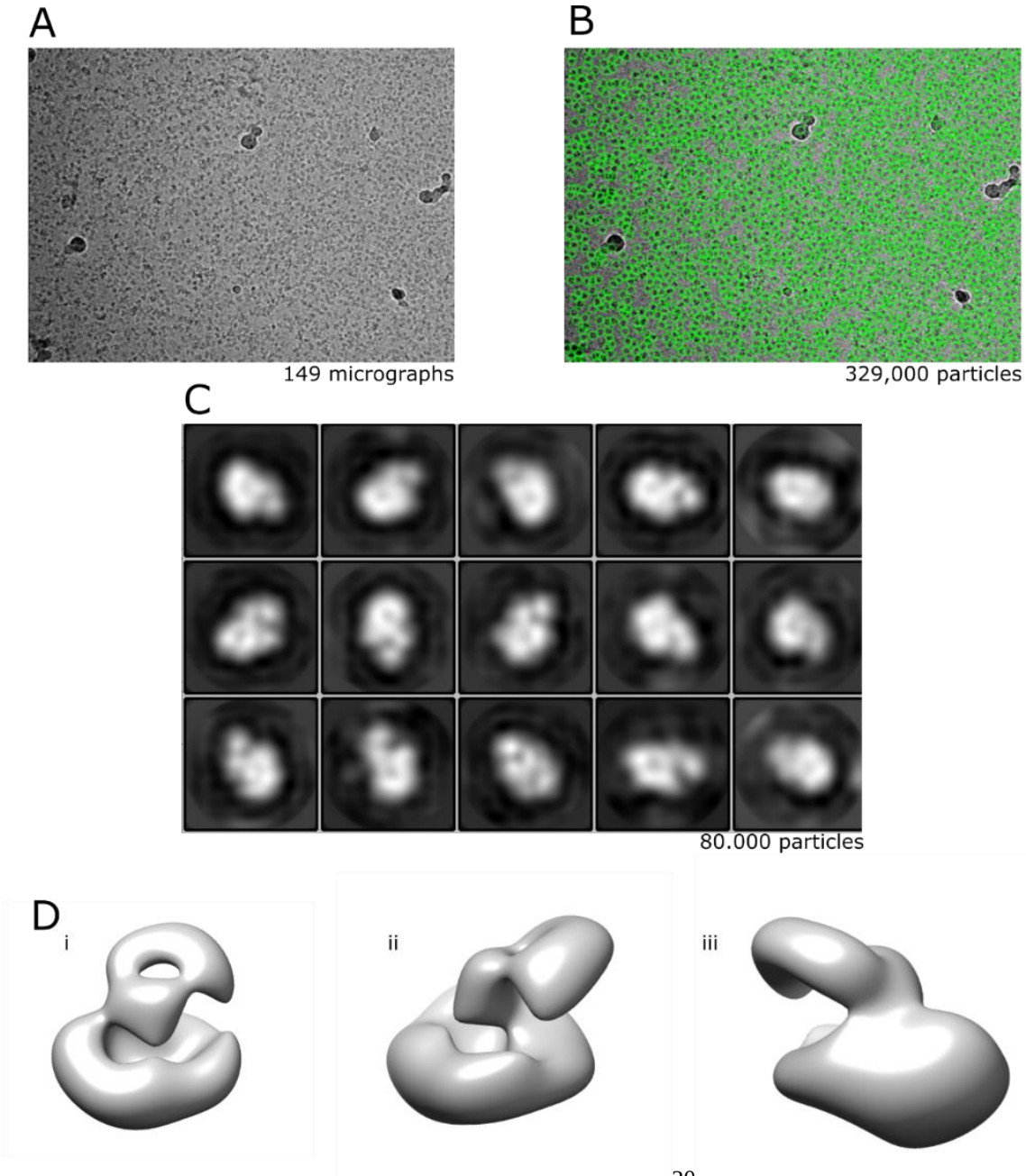


Figure 31: Model of uncrosslinked Hfq and ChiX subcomplex. A) Micrograph collected of grid made in figure 29. B) Particles picked by Relion autopicking software. Each green circle represents one particle. Some dirt and overlapping particles have been erroneously picked, which were manually removed. C) Final 2D classification of particles. During 2D classification, poorly aligning particles were discarded, ending up with around 80,000 particles. In the 2D classification, the classifications were found too small to be the full supercomplex expected. D) 3D recontruction of the final particles.

To further stabilize the supercomplex, I used glutaraldehyde to crosslink the individual components together. However, the crosslinked complex was found aggregated on the carbon of the copper quantifoil grids and therefore not able to be imaged. Through various trials and errors, I successfully imaged the crosslinked supercomplex on gold grids. The process of sample preparation is summarised in Figure 32.

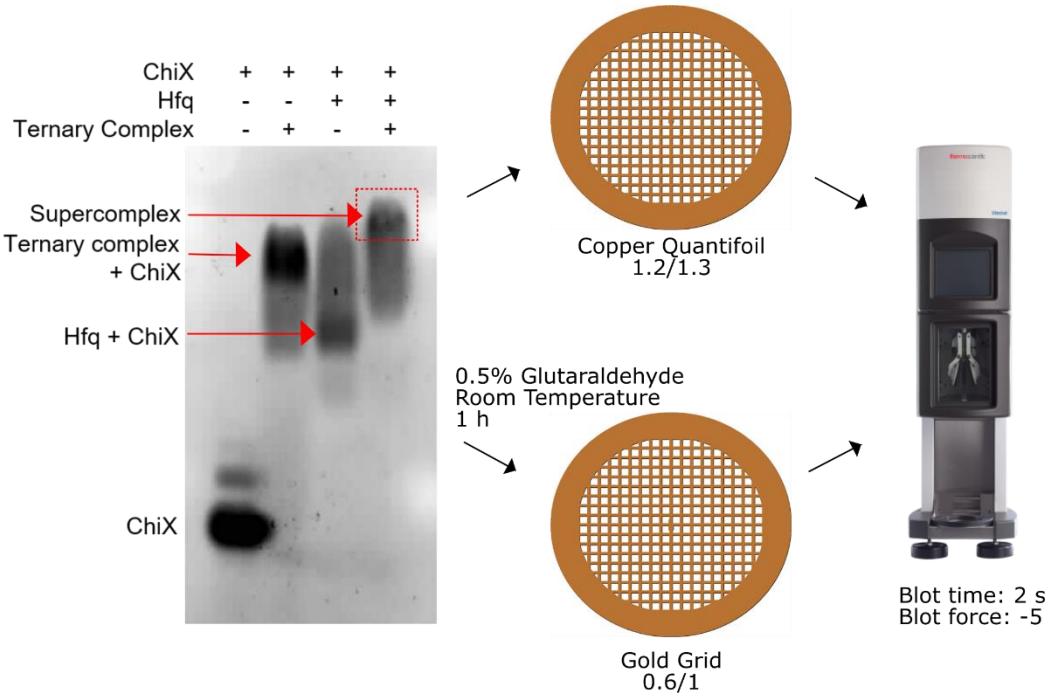


Figure 32: Sample preparation of uncrosslinked and crosslinked supercomplex. Uncrosslinked Supercomplex was formed with Ternary complex (RhlB E166Q, RNase E 603-850, enolase), ChiX, and Hfq. Supercomplex was added to copper quantifoil grids and blotted with the vitrobot and frozen in liquid ethane. Crosslinked Supercomplex was crosslinked with 0.5% glutaraldehyde at 25 °C for 1 h. Crosslinking was quenched with 200 mM of glyine pH 7.5. Crosslinked proteins were put onto a gold grid and plunge frozen in liquid ethane. Frozen grids were stored in liquid nitrogen.

Data collection was performed in the Cambridge Biochemistry department, with a Titan Krios fitted with a K2 detector at a pixel size of 1.43 Å per pixel. 1570 micrographs were collected over a 48-hour period using the EPU software. Data pre-processing (motion correction and CTF correction) and automated particle picking were performed using WARP (Tegunov & Cramer, 2018). The picked particles were imported into Relion and 2D classification was performed. Through the 2D classification, three different sizes of particles were identified. Smaller particles were separated from the Medium particles which were also separated from the Large particles. After separation, the three different particles were re-classified using 2D as well as 3D classification. Poorly aligned classes were removed in order to removing bias. The particle number distribution however was uneven, with the largest group being the Medium sized particles. 2D classification was continued independently for each of the groups, removing ill-fitting particles for each class. 3D classification was performed for the Small sized particles (Figure 33E), which are the focus of this thesis. 8271 clean Smaller particles were obtained in total. The 2D classifications showed again the characteristic hexagonal donut shape of Hfq (Figure 33E), which led me to believe that this is the same subcomplex which was previously seen. The sizes of the 2D classifications were comparable, however, internal structural features were resolved better than with the previous dataset. The set of clean Small particles were subsequently reconstructed in 3D using Relion. The 3D reconstruction was calculated to be 25Å in resolution (GS-FSC) (Figure 33).

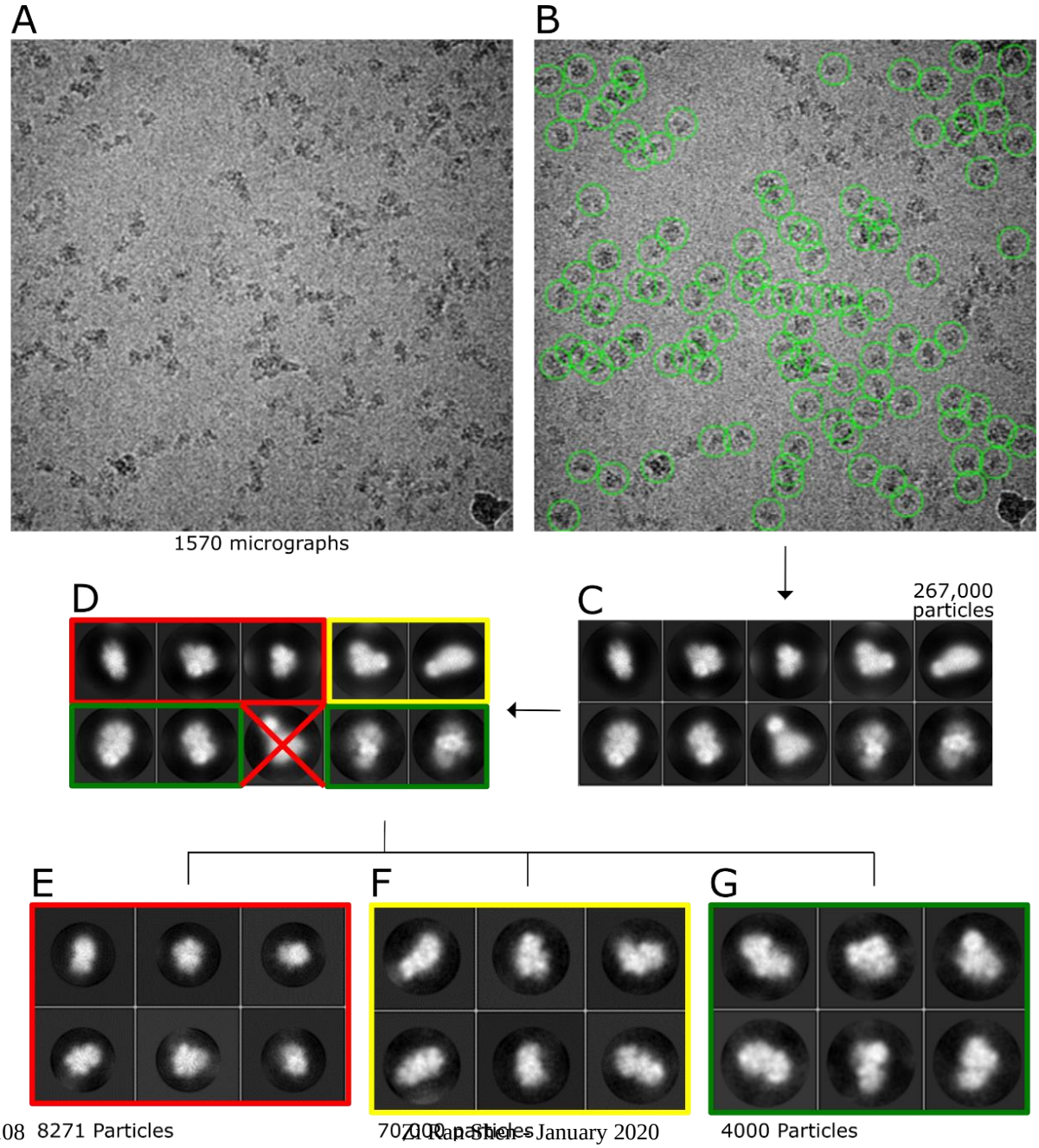


Figure 33: Data processing of supercomplex micrographs. A) Obtained supercomplex micrographs from the Titan Krios intrument using a K2 detector. B) Particles outlined in green circles. C) 2D alignment of all particles. Multiple sizes of particles can be shown in the alignment, which is color coded in subfigure D. The red outline shows the Small subsection of particles, while the yellow showed the Medium subsection. The green outline showed the Large subsection of particles. All particles of different sizes were separated into different particle files to prevent misalignment. All separated classes were re-classified in figures E, F, and G. The corresponding numbers of particles are shown below the classifications. Group E is the main topic of this thesis.

Molecular modelling was performed in Chimera (Pettersen et al, 2004). ChiX contains, from 5'-3', a stem-loop, the Hfq distal face binding site, the seed region, another stem loop, and finally the Hfq proximal face binding site (Figure 35). The two Hfq binding sites were modelled first, according to previously reported structures of Hfq bound to RNA. The distal face binding site (Figure 34Ci) was modelled after the structure of Hfq bound to poly-A RNA (Link et al, 2009) while the proximal face binding site was modelled after the structure of Hfq bound to RydC (Dimastrogiovanni et al, 2015). The ChiX stem-loop structures were modelled using RNAComposer (Biesiada et al, 2016) with secondary structure information from Figueroa-Bossi et al, 2009. The flexible regions were modelled by hand.

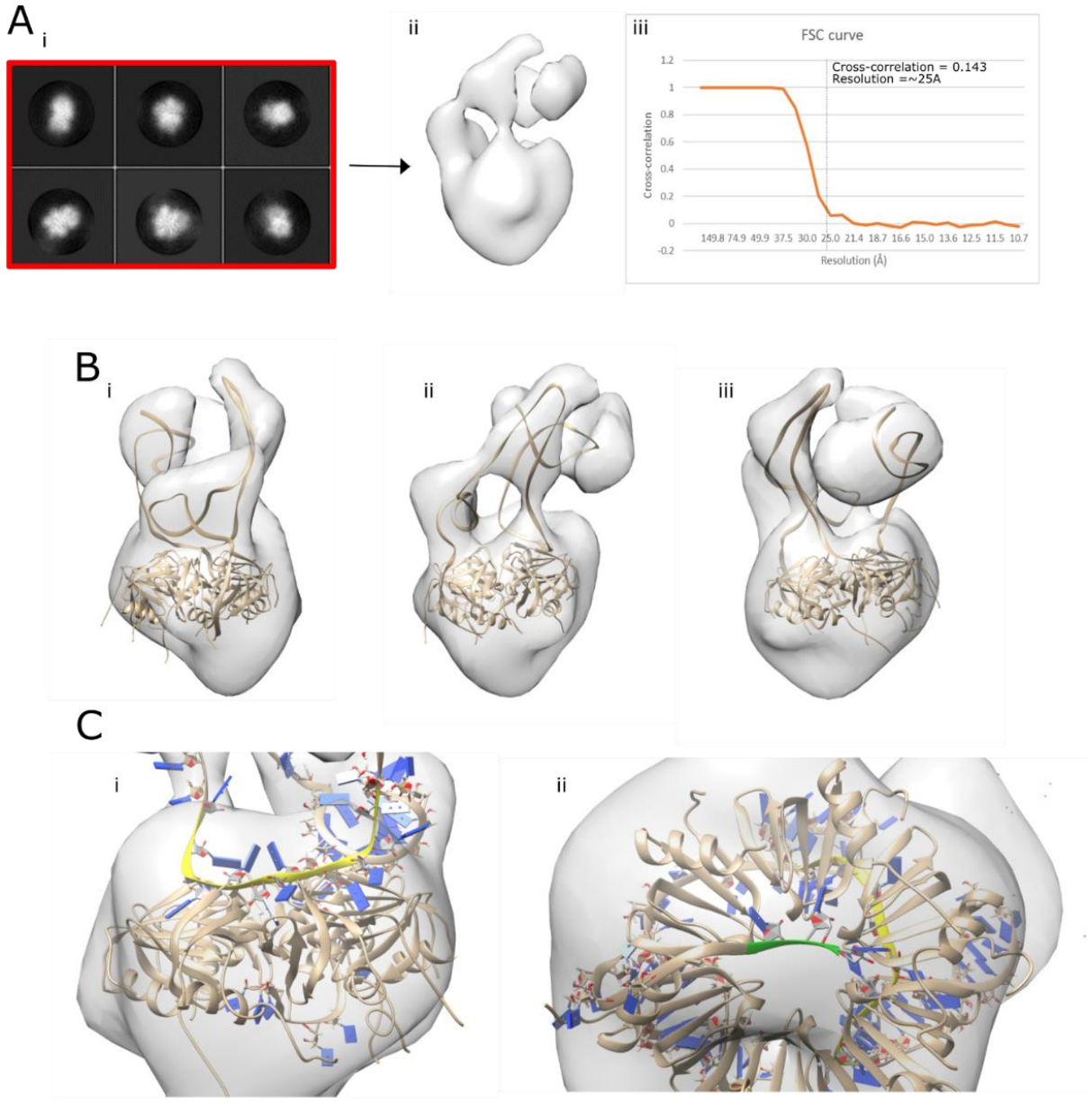


Figure 34: Putative model of ChiX bound to Hfq. A) The 2D classifications (i) show a hexagonal floret which may potentially be Hfq. As the 3D reconstruction (ii) is low in resolution (iii), the exact composition of the particle is indeterminate. Nevertheless, as the 3D reconstruction was real - the same data was reclassified multiple times and all maps were identical - and the density was too large for an enolase dimer, I attempted to create a model of Hfq bound to ChiX based on the map. However, the possibility that this density represents a different subassembly should be considered for future analyses. B) Hfq bound to ChiX fitted to the volume. ChiX Has two binding sites for Hfq and two stem-loops, which aided in the formation of the model. Note that there is a substantial amount of density on the proximal side of Hfq (bottom of image) which is unfitted. C) The two Hfq binding segments of ChiX highlighted in yellow and green. The yellow segment binds to the distal face of Hfq (modelled after the crystal structure deposited in PDB (accession code 2GIB) by Link et al in 2009) while the green segment binds to the proximal face (modelled after the crystal structure (PDB accession code 4V2S) deposited by Dimastrogiovanni et al in 2015).

Figure 34 shows the model of Hfq bound to ChiX, depicted in ribbon representation. As the resolution of the map is low, only the general shape of the RNA can be inferred. Interactions between the RNA and Hfq, as well as interactions between ChiX's flexible regions cannot be identified, but a likely interaction mode can be proposed for the proximal and distal face binding sites of ChiX with the RNA binding sites on Hfq. The bulk of the RNA appears to be concentrated on the distal face of Hfq and not evenly wrapped over the surface of the chaperone. Note as well that, on the proximal face of Hfq, there remains a feature in the map which is not filled by ChiX or Hfq. These properties will be further discussed in the following section.

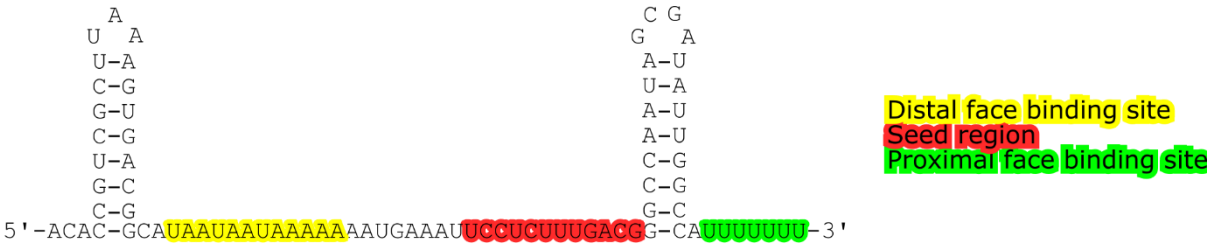


Figure 35: Schematic of ChiX and its binding sites of interest

4.3 Discussion

Though the map of Hfq bound to ChiX is limited by its low resolution, the information from the shape allowed a potential mode of interaction to be proposed. The model of Hfq bound to ChiX was derived in part from pre-existing structural information of Hfq bound to other RNAs as well as the secondary structure and function of each ChiX segment. The molecular model of ChiX bound to Hfq shows a striking feature, which is that the mass of RNA is likely

to be concentrated on the distal face of Hfq rather than evenly spread between the proximal and distal faces, especially since there is a stem-loop directly connected to the proximal face binding site. In the model above, more than 90% of ChiX mass appears to be positioned on the distal face of the bound Hfq. This configuration may serve as a sensor to target RNA binding, or it may confer protection to the flexible seed region (Figure 36).

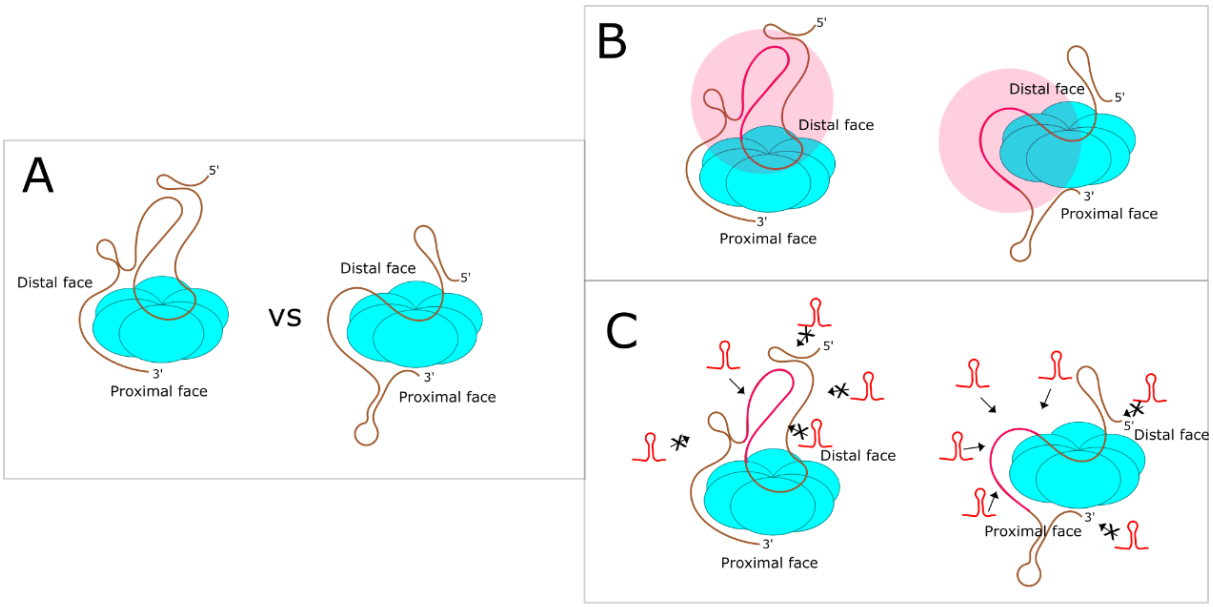


Figure 36: Cartoon schematic of ChiX bound to Hfq. A) The two representations considered in the discussion. The cartoon on the left is modelled after the experimentally determined map while the cartoon on the right is a more "intuitive" interpretation that serves as a foil to explain the advantages of the experimentally determined model. B) Figure showing the immediate area which may be affected by the seed region upon binding to its RNA target. The red segment of the RNA represents the seed region while the pink circle represents the area which may be contacted by the seed region. On the left, most of the RNA is in contact, and may thus be immediately affected upon any structural changes to the seed region. In contrast, the seed region in the model on the right only directly affects the two RNA segments covantely attached to it. C) Figure showing the accessibility of the seed region. The model on the left has a limited angle at which RNA may approach the seed region, whereas the seed region in the model on the right may be approached from a wider angle.

As a sensor, the organization of ChiX on Hfq enables the majority of the sRNA to contact with the seed region. In the model shown above, the seed region seems to interact with multiple elements of ChiX all at the same time. It is bound directly to both the 3' stem-loop as well as the distal face binding site. The seed region is shown to potentially contact the 5' stem-

loop as well. Through this configuration, the seed region can directly affect more than half of ChiX excluding itself (Figure 36B). If ChiX was distributed evenly across both the proximal and distal face of Hfq, the seed region would need to straddle the rim, and thus can only affect the nucleotides it is directly bound to (Figure 36B). This configuration would limit the efficacy of the communication between the seed region and other regions of ChiX. Thus, crosstalk between the seed region and the 5' or 3' ends of ChiX may be significantly greater if the majority of the RNA is concentrated at the distal face of Hfq. An additional feature of this geometry is that it may confer "protection" to the flexible seed region. As the seed region is fairly long, if it is exposed it can form unwanted non-specific interactions with other RNA molecules. Surrounding the seed region with stem-loops limits the points of access on which an RNA can base pair, thus allowing only for the more robust interactions by partners of ChiX (Figure 36C). This hypothesis is in part supported by the finding that most class II sRNAs (RNAs which bind to both sides of Hfq) have more specific targets than class I sRNAs, which bind a larger variety of RNAs (Schu et al, 2015). This arrangement of ChiX bound to Hfq may have larger implications with regards to the geometry of all class II sRNAs, as most class II sRNAs have only a small pool of binding partners and thus may require protection in the seed region.

In the Hfq ChiX map, there is prominent density on the proximal face of Hfq which remains unaccounted for. This density may encompass the flexible rim region of Hfq. Most crystal structures of Hfq omit the C-terminal flexible rim region, which has been recently shown to play a role in RNA interactions (Santiago-Frangos et al, 2019; Santiago-Frangos & Woodson, 2018). The proximal density may be a result of the movement of the C-terminal flexible regions sweeping over the proximal face binding site. As the flexible regions of Hfq serve to repel

RNA, they may be protecting the proximal binding site from other class I sRNAs from competing with the already bound ChiX. Class II sRNAs have been shown to be able to out-compete class I sRNAs for Hfq binding. Thus, this protection mechanism may favour class II sRNAs over class I sRNAs.

4.4 Methods

4.4.1 Hfq purification

Competent BL21 cells were transformed with the plasmid that encodes Hfq conferring ampicillin resistance obtained from Dr. Heather Bruce. Transformed cells were plated on ampicillin containing (30 μg/mL) agar plates and incubated at 37 °C overnight (16 h). Starter culture was made by inoculating 20 mL of LB media with 30 μg/mL of ampicillin with one colony of transformed cells. Inoculated media was grown overnight (16 h) at 37 °C with aeration. 2 x 1 L of LB media with 30 μg/mL of ampicillin was inoculated with 2 x 10 mL of starter culture. Inoculated 1 L cultures were grown at 37 °C until cells reached mid-log phase (OD_{600} = 0.6). 1 mM IPTG was added to 1 L cultures to induce protein expression. Cultures were harvested 5 h after expression induction via centrifugation at 4,000 x g for 20 min. Cell pellets were stored in -80 °C prior to lysis. On the day of purification, frozen cells were thawed in Hfq lysis buffer (for each gram of dry cell mass, 3-5 mL of lysis buffer was added). Thawed cells were lysed via high pressure homogenization using an Avestin Emulsiflex C5. Thawed cells were passed through the homogenizer at 5000 psi four to six times, or until the DNA was sheared. Cell debris was separated from soluble proteins by centrifugation at 30,000 x g for 30 min. Clarified cell lysate was heated to 85 °C for 60 min. Heated cell lysate was centrifuged at

30,000 x g for 30 min. Clarified lysate post-heat treatment was diluted in a 1:1 ratio with 4 M ammonium sulfate. Diluted lysate was incubated on ice or at 4 °C. Diluted lysate was then centrifuged at 30,000 x g for 30 min. Supernatant was collected. Meanwhile, a HiTrap Butyl HP column was equilibrated in ddH_2O (4-5 column volumes) and Hfq HIC binding buffer (table 5, 4-5 column volumes). Processed cell lysate was then loaded onto an equilibrated column. The loaded Butyl HP column was then washed with 4 column volumes (CV) of 20 % Hfq HIC elution buffer (300 mM ammonium sulfate total) (Table 5), or until A_{280} reached baseline. Hfq was eluted with a gradient of 20-10 % Hfq HIC elution buffer over 50 mL. Hfq was collected in 2 mL fractions. Relevant fractions were pooled and concentrated. Concentrated fractions were loaded onto equilibrated S200 column. 0.3 mL fractions were collected. Peak fractions were visualized on an SDS-PAGE gel. Fractions containing clean Hfq were concentrated and stored in -80 °C.

Table 5: List of buffers used for Hfq purification.

Buffer Name	Buffer composition
Hfq lysis buffer	Tris pH 7.5 (50 mM), NaCl (250 mM), KCl (100 mM), $MgCl_2$ (10 mM), β-mercaptoethanol (50 μM), 1 x cOmplete™, EDTA-free Protease Inhibitor Cocktail (Roche) per 50 mL
Hfq HIC binding buffer	Tris pH 8.0 (50 mM), Ammonium sulfate (1.5 M), β-mercaptoethanol (50 μM)
Hfq HIC elution buffer	Tris pH 8.0 (50 mM), β-mercaptoethanol (50 μM)
Hfq S200 buffer	Tris pH 7.5 (50 mM), NaCl (200 mM), KCl (250 mM), $MgCl_2$ (10 mM) β-mercaptoethanol (50 μM), glycerol (10 % v/v)

4.4.2 ChiX IVT

A plasmid containing ChiX and primers required to amplify the DNA template was kindly provided by Dr. Wenxia Sun. PCR mix was made with 1 x Phire reaction buffer, 10 ng template DNA, 200 μM dNTPs, 0.5 μM of both primers received, and 1 μL Phire hot start polymerase. The template DNA used was the ChiX plasmid received above. PCR temperature cycle was programmed as follows: Initial denaturation at 98 °C for 2 min, followed by 32 cycles of denaturation at 98 °C for 5 s, annealing at 60 °C for 5 s, extending at 72 °C for 1 min, and then a final extension step at 72 °C for 10 min. Amplified ChiX PCR template DNA was purified with a Qiagen PCR purification kit. Purified IVT template DNA was used to transcribe ChiX RNA. In-vitro transcription (IVT) reaction mix was made with 1 x reaction buffer (50 mM Tris pH 7.5, 10 mM NaCl, 5 mM $MgCl_2$, 200 μM spermidine), 100 μM rNTPs, 3 μg of T7 polymerase, 1 μg of template DNA in 4 mL of RNase free H_2O. IVT reaction mix was incubated at 37 °C for 4 h. RNase Free DNase I was then added to the reaction mix and incubated for 30 min at 37 °C. All reactions were stopped with the addition of RNase free 10 mM EDTA. Stopped reactions were run on a 10 % urea sequencing gel for 5 h. The gel was illuminated with short wavelength UV light to identify the relevant RNA band, which was then excised. RNA was electroeluted from the gel and precipitated with 300 mM sodium acetate pH 5.3 and 3 volumes of ethanol in -20 °C overnight (16 h). Precipitated RNA was harvested by centrifugation. The rNA pellet was dried and stored in -20 °C.

4.4.3 Supercomplex formation and crosslinking

The supercomplex consisting of RNase E 603-850, RhlB E166Q, enolase, Hfq, and ChiX was made by combining all elements in an equimolar ratio. Combined supercomplex was

purified by size exclusion chromatography. Formation of the supercomplex was verified by EMSA. Alternatively, after combining all elements of the supercomplex, glutaraldehyde crosslinking was performed with 0.5 % glutaraldehyde at 25 °C for 1 h. Crosslinking was quenched with 200 mM of glycine pH 7.5. The products of the quenched reaction were purified using the size exclusion column.

4.4.4 Grid preparation

Copper quantifoil grids were first glow-discharged for 1 minute at 25 mA. 3 μL of purified supercomplex (not crosslinked) at a concentration of 0.1 A_{280} was then added to copper quantifoil grids. Excess liquid was blotted with a blot force of -5 and blot time of 2.0 s. Loaded grids were frozen using a Vitrobot. Frozen grids were stored in liquid nitrogen until data was collected. Gold grids were also glow-discharged for 1 min at 25 mA. 3 μL of purified supercomplex (crosslinked) at a concentration of 0.1 A_{280} was then added to the gold grids. Excess liquid was blotted with a blot force of -5 and blot time of 2.0 s. Loaded grid was frozen using a Vitrobot. Gold grids were similarly stored in liquid nitrogen until data was collected.

4.4.5 Data collection

Images of the copper quantifoil grids were collected in a Titan Krios instrument fitted with a Falcon II detector. Images were collected with EPU software over 48 h. 149 micrographs were collected. Gold grid images were collected in a Titan Krios fitted with a K2 summit detector. Images were collected with EPU software over 48 h.

4.4.6 Data processing

For images of uncrosslinked proteins, images were corrected for both beam induced motion and stage motion using MotionCor2 (Zheng et al, 2017). Following motion correction, images were corrected for defocus and lens induced aberrations using Gctf (Zhang, 2016). 2000 particles were picked from corrected micrographs and 2D classifications were made using Relion 2.1 (Scheres, 2012). Suitable 2D classifications were used as templates for particle autopicking. Autopicked particles were sorted and junk particles (particles which were erroneously picked, particles which are clearly not of proteins, or particles with poor resolution) were discarded. The remaining particles were classified in 2D classification. Poorly aligning particles were discarded. 2D classification and discarding particles were performed 10 times to obtain a clean and well aligning set of particles. Ab-initio 3D model generation was performed with Relion 3.0. The initial model created was used to classify the remaining particles into five 3D classes.

For images of crosslinked proteins, images were corrected for both beam induced motion and stage motion, defocus and lens induced aberrations using Warp (Tegunov & Cramer, 2018). Particles were also autopicked from the micrographs using Warp (Tegunov & Cramer, 2018). As Warp is highly selective with the particles picked, very few particles were junk. All picked particles were classified in 2D classification. Poorly aligning particles were discarded. 2D classification and discarding particles were performed 3 times to obtain a clean and well aligning set of particles. Three different sizes of particles were noted, and particles corresponding to each particle size was separated into individual coordinate files. Particles of each size was aligned in 2D with other particles of similar size, while discarding poorly aligning particles. This cycle was repeated around 3 times for each particle size. *Ab-initio* 3D model

generation was performed with Relion 3.0 for each size. The initial model created was used to classify the remaining particles into five 3D classes per size.

5 TOWARD UNDERSTANDING HOW SMALL RNAS ARE RECOGNISED BY RNASE E IN THE DEGRADOSOME

5.1 Introduction

The small domain of RNase E forms a self-complementary interface that contributes to the oligomerization of the enzyme as a dimer-of-dimers. Recently, the small domain has also been shown to form an RNA interacting surface together with the RNase H domain of RNase E (Bandyra et al, 2018). This surface can bind to a duplex region of the small RNA RprA (Figure 37). With the discovery of this novel interaction site, more questions have been raised about the RNA binding capacity of the small domain and therefore RNase E. When the putative RNA binding residues on the small domain were mutated, a marked decrease in efficacy of RNA degradation by the direct entry mode of action was observed. When the 5'-triphosphate of the RNA substrate was switched to a 5'-monophosphate, the efficacy was not strongly affected. These results suggested that the small domain is not only a structural domain that organises quaternary structure, but is also a recognition domain responsible for mediating interactions with secondary structure in substrates. Through this mode of recognition, the small domain potentially contributes to a cleavage mechanism which bypasses the requirement to recognise the 5'-ends of RNA substrates.

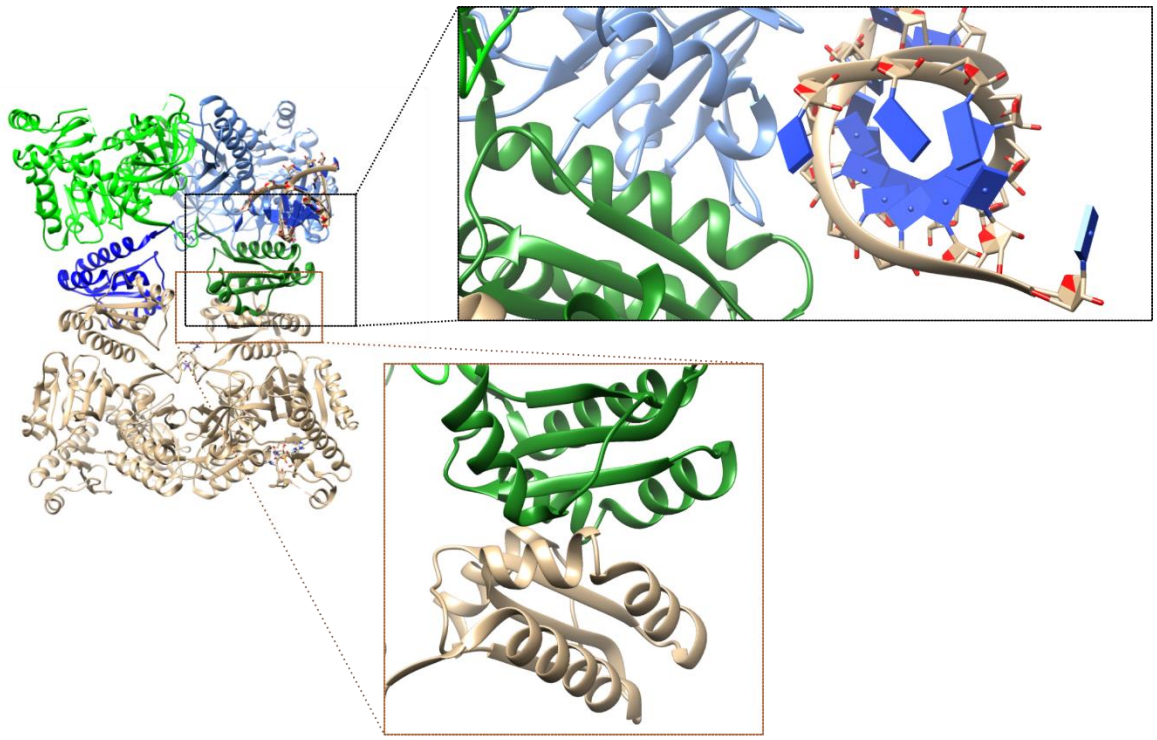


Figure 37: RprA bound to RNase E (Bandyra et al, 2018). RprA is seen in the image as binding to two different monomers, one colored green and one colored blue. It contacts the blue monomer through the DNase I domain and the green monomer through the small domain. The RNase E small domain dimerizes through the two helicies connected to each other. The two helices contact another helix on the other dimer as well as a beta sheet.

Recently, a report revealed the sequence similarity and structural congruence between the KH domains and the small domain of RNase E (Pereira & Lupas, 2018). Using a Hidden Markov analysis of the amino acid sequence patterns (jackhmmer), the small domains of RNase E/G were reclassified as a novel type of KH domain. The structural similarity between the KH domains and the RNase E small domain is evident when compared (Figure 38). Although both

KH domains and the RNase E small domain have a different organization, they share a similar 3D architecture.

KH domains contain 3 alpha helices and 3 β-strands (Figure 38A). α' and the β' strand in the KH-1 and KH-2 domains are not previously known to bind RNA, whereas α1 and α2 contain a conserved motif IGxGxxIK/R that is used to bind nucleotides. In some proteins, α' and β' have been seen to be involved in dimerization (Lewis et al, 1999; Teplova et al, 2011). Some KH domains dimerize to create a larger binding site to bind a longer chain of nucleotides (Jia et al, 2010). Canonically, KH domains' nucleotide binding interfaces can only accommodate single-stranded oligonucleotides (Valverde et al, 2008). Though a single KH domain can only accommodate one strand, multiple KH domains can associate with each other for different modes of binding. Indeed, many KH domain containing proteins have multiple copies of the domain (Valverde et al, 2008). Not only that, KH domains' ability to oligomerize suggests that the mode of nucleotide binding by KH domains may be highly diverse (Lewis et al, 1999; Valverde et al, 2007).

Though the RNase E small domain shows a similar architecture, the apparent functions of the helices and sheets are different. The loop between α' and β' was shown to bind RNA (Figure 38C), and α1 and α2 are involved in dimerization (Figure 37) (Bandyra et al, 2018). In this chapter, I explore the current available structures of KH domains and compare them to the small domain of RNase E. Through this analysis, I hope to further understand the extent to which the small domain of RNase E can bind RNA and also the extent to which RNA-Hfq complexes may associate with RNase E through shared interactions with secondary structural elements and other features of the RNA.

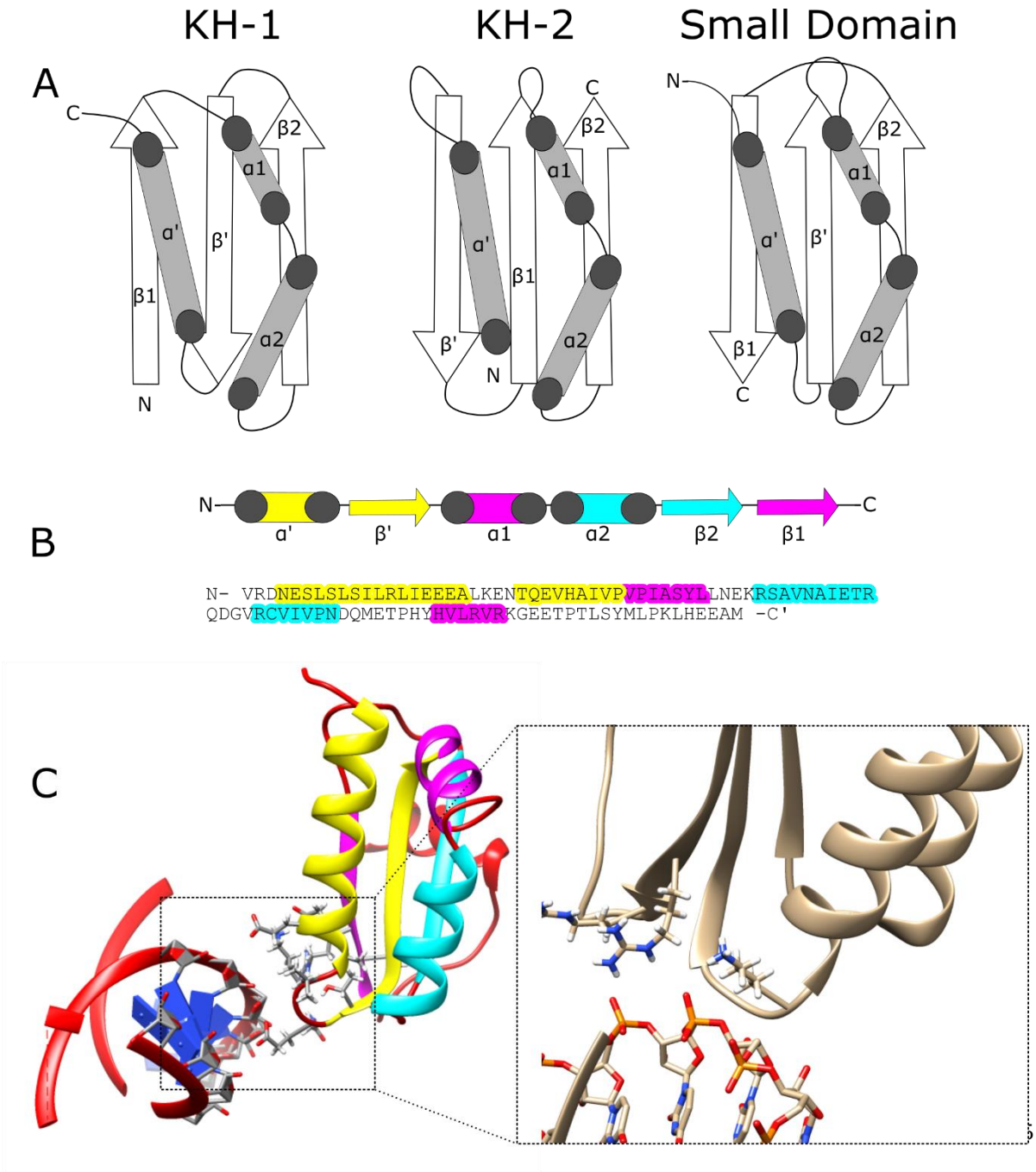


Figure 38: Organization of RNase E small domain as a KH domain. A) Cartoon of 3D organization of RNase E small domain compared to KH-1 and KH-2 domains. Though the organization of the structures are different, the overall 3D fold is similar. B) Sequence of RNase E 413-500 colored by chain. The KH domain chains are organized by color. Chains α' and β' are colored yellow, chains α1 and β1 are colored magenta, while chains α2 and β2 are colored cyan. C) The structure of RNase E small domain bound to RNA. Two residues seem to be binding to the RNA backbone. One is a lysine located in a loop between α' and β', while another is an arginine located on β1.

5.2 Results

The amino acid sequence of the small domain of RNase E (residues 413-531) was used in a jackhmmer search (Potter et al, 2018). jackhmmer is an iterative search tool based on the software hmmer, which uses hidden Markov Models to search all sequence databases accessible by UniProtKB. The small domain of RNase E was used as a search query, which yielded 1690 hits in the first iteration. These hits were then used as the query for the next iteration of jackhmmer. An iterative search was performed until the number of hits stabilized at 33695 sequences (Figure 39). In each iteration, the e-value for each hit was measured. Hits above the threshold e-value (0.01) were discarded from the data. Stabilization of the number of hits was described as +/- 500 hits for 3 iterations. Out of all hits, only 209 hits did not map to the KH domain. The rest of the hits all aligned to different kinds of KH domain across all kingdoms of life (Figure 39).

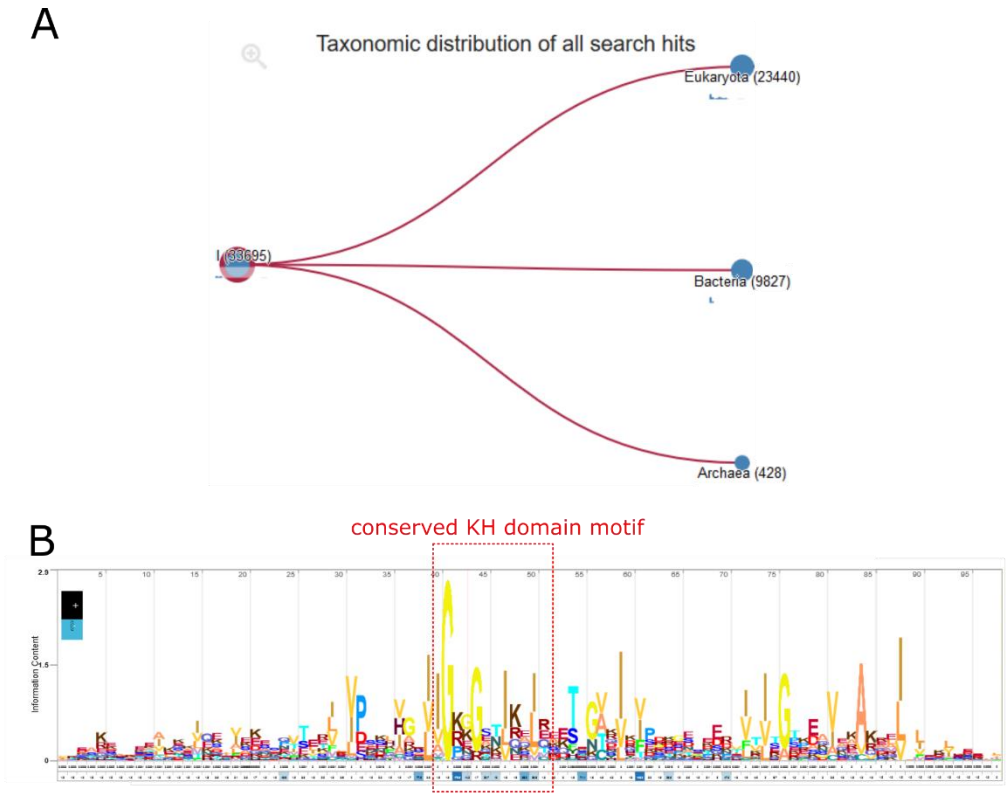


Figure 39: jackhmmer results following search with query of RNase E 413-531. A) Taxonomic tree showing distribution of hits across the three kingdoms of life. Eukaryota has the most hits at 23440, followed by Bacteria with 9827 hits, then finally Archaea with 428 hits. B) HMM logo generated by the hits. The most striking conservation is the glycine at position 41, which is almost 100% conserved. Surrounding the glycine is the KH domain motif, IG--G--IK/R.

Following sequence analysis, I decided to assess how different KH domains may bind to RNA, and whether or not previously described KH domains have been shown to bind RNA in a similar way as the RNase E small domain (Figure 37). I utilized the databases Interpro, Pfam, SMART, Superfamily, and PROSITE to gather structural information on previously described KH domains. In total, 926 non-redundant PDB codes were obtained across all five databases. In those 926 structures, 677 structures contained bound RNA. In the structures

containing KH domains with RNA bound, 412 KH domains were shown to potentially have the same RNA binding arrangement as the small domain of RNase E. An additional 15 KH domain structures associated with RNA were found in the same geometry as the RNase E small domain bound to RNA, but either lack of sidechain information or poor resolution deemed the binding probable, though inconclusive. All KH domains which showed interactions with RNA in a similar fashion to the RNase E small domain were found in the ribosome. However, no mitochondrial ribosomes surveyed contained this binding geometry. Notably, in all structures surveyed, no KH domains showed no dimers which matched the dimerization conformation seen in the RNase E small domain.

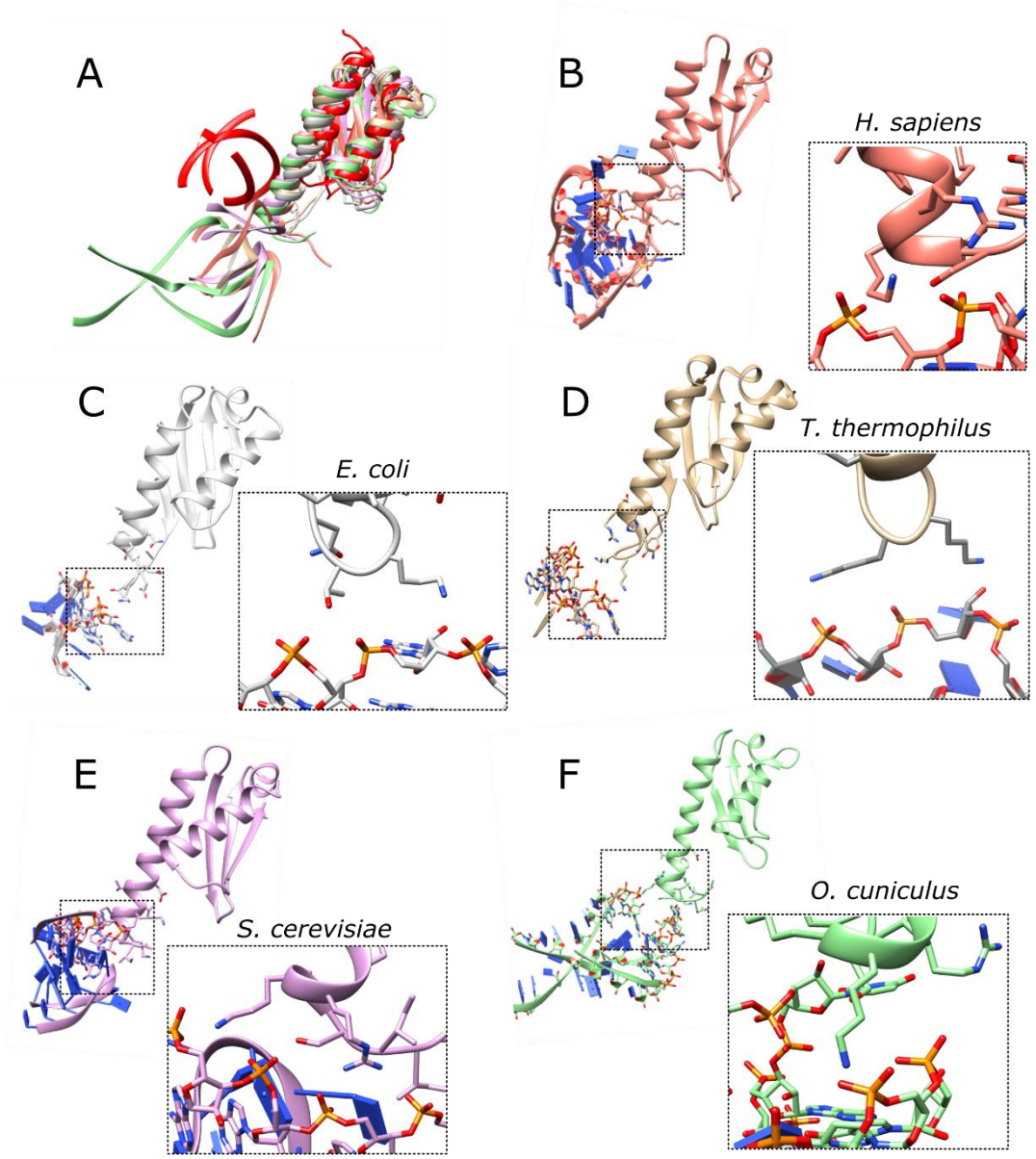


Figure 40: A selection of different KH domains from different species binding to RNA. A) Alignment of RNase E small domain in red to all structures shown in this figure. The body of the KH domains align closely, while the RNA bound to RNase E small domain is slightly higher than ribosomal RNA. However, the helix which binds to RNA is slightly shorter in RNase E than in the ribosomal proteins. Subfigures B-F show the details of RNA binding in each structure. There are two modes of binding which can be seen, one exhibited by the prokaryotes (C and D) and one exhibited by the eukaryotes (B, E and F). In the eukaryotic ribosomes, a phosphate in the RNA backbone is contacted by a lysine and an arginine in the N-terminal of the α' helix, which is distant from the α1 helix. In the prokaryotic ribosomes, the phosphate backbone is contacted by a lysine and a lysine or threonine. These residues are found in a loop furthest away from the α1 helix.

Comparing the RNA binding mode of RNase E small domain (Figure 38C) and the ribosomal KH domain RNA binding mode (Figure 40, B-F), similarities can be seen. In the RNase E small domain, one lysine and one arginine can be seen binding to the RNA phosphate backbone. In prokaryotic ribosomes, there is also one conserved lysine which binds the phosphate backbone located in the furthest loop away from α1 next to α'. In eukaryotic ribosomes, two residues are responsible for binding the RNA phosphate backbone, one lysine and one arginine. Though the two residues are located in the helix and not in a loop, their position is the helix is very close to the terminus. Geometrically, the α' helix in the RNase E is shorter than the α' helix in the ribosome, which may account for the differing orientations of the RNA in Figure 40A.

Sequence analyses using jackhmmer were performed for other domains of RNase E, namely the RNase H and DNase I domains (Figure 5). The RNase H domain was queried with RNase E 1-39 and 211-280 combined iteratively until no new sequences were added to the search. A total of 7040 sequences were found, most of which were aligned with an RNase E/G like domain in prokaryotes. 42 sequences were aligned with either no domain architecture or S1 domain architecture (based on Pfam domains), suggesting a close relationship between the prokaryotic S1 domain and RNase H like domain. The DNase I domain search was performed with RNase E 281-400 as the query. The search was performed iteratively until no new sequences were added to the results. The search returned 7666 sequences, most of which were also aligned with an RNase E/G like domain architecture. Only 144 sequences were not aligned with any domain architecture. These results suggest that the RNase H and DNase I domains are

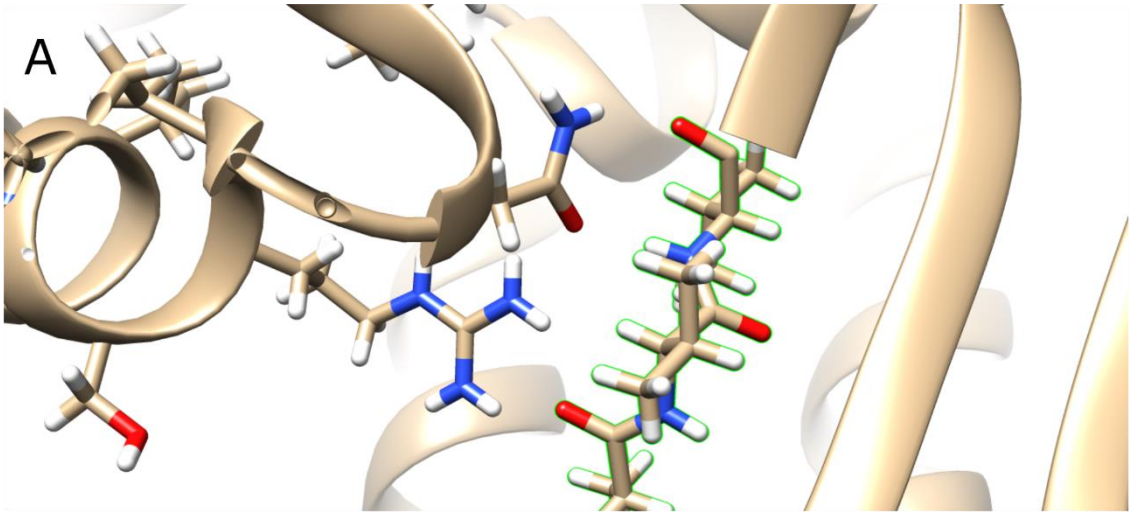
not distantly related to any other domains, and that these domain families may have more defined and specific roles than the KH domain family.

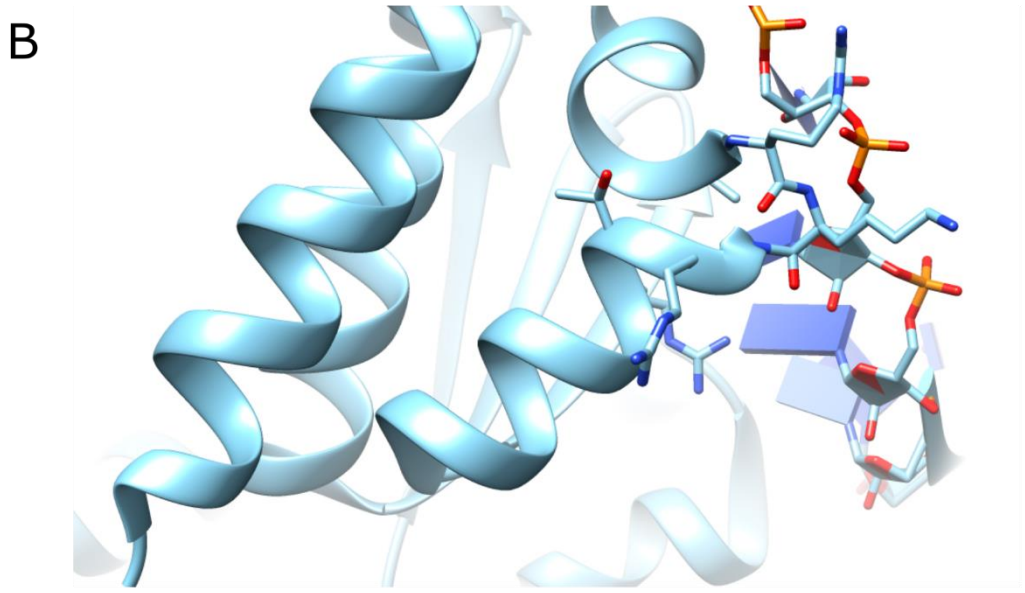
5.3 Discussion

Recently, the RNase E small domain has been shown to be a third class of KH domains (Pereira & Lupas, 2018). Though lacking the IGxGxxIK/R KH motif, the RNase E small domain retains the geometry of the KH domain. As the KH motif is used to bind single stranded RNA/DNA, the RNase E small domain may not be able to bind RNA in the same manner. Indeed, the α1 and α2 helices — where the KH motif would reside — in the small domain (Figure 37) are involved in self-association and are therefore inaccessible for RNA binding. Though the KH motif is inaccessible for RNA binding, the α' loop has been shown as a conserved RNA binding site in all three different classes of KH domains (Figure 40). Association with RNA by the α' loop must be limited as only two residues are seen to be involved in RNA binding. In addition, these residues are shown to interact with the backbone, supporting non-specific binding. These observations support the previously obtained structure of RNase E bound to RprA, and the potential involvement of the small domain in the direct entry pathway of RNase E cleavage.

One notable result found through this search is the lack of KH domain dimers which adopt a similar conformation to the RNase E small domain dimers. The residues involved in KH domain dimers are generally found in α', which was used in RNase E to bind RNA (Lewis et al, 1999; Teplova et al, 2011). The residues of the RNase E small domain involved in dimerization are along the α1 and α2 helices, which are involved in RNA binding in canonical

KH domains. This characteristic may be related to the absence of the notable IGxGxxIK/R motif in the RNase E small domain, as the KH motif is used for RNA binding and generally appears in the α1 and α2 helices (Figure 41). RNase E instead uses these binding residues to dimerize with the β2 strand of the other dimer (Figure 41). These results suggest the potential of the RNase E small domain to have diverged from the other KH domains early on its evolution, as it only retains the overall 3D structure and the RNA binding site between α' and β'.





C

```
PH1566      ALPRVRGRIIGRKGRTRQIIEEMSG 25
RNaseE      -VPI-ASYLLNEKRSAVNAIETRQ- 22
             :*  . : : * . : **  .
```


Figure 41: Comparison of substrate binding by the α1 and α2 helices of RNase E small domain and KH domain. A) Dimerization surface of RNase E small domain. One of the dimers (left) is associating with the backbone of the other dimer (right) in the α2 helix. Other residues of the α1 and α2 helices such as the asparagine are also involved in associations with the dimeric partner of the RNase E small domain. B) KH domain binding to RNA. The arginine and lysine are interacting with the phosphate backbone of the RNA. C) sequence alignment between the KH domain of PH1566 and the RNase E small domain. The highlighted sequence is the conserved IGxxGxxIK/R motif found in the KH family. Most of the highly conserved residues of this motifs are present in PH1566, but absent in RNase E

5.4 Materials and Methods

The RNase E small domain was used as an initial query in the iterative Hidden Markov Model search jackhmmer (Potter et al, 2018). Results were used as new queries until search converged. E-values were set at 0.01 and sequences were scored using the BLOSUM62 matrix. The RNase E domains of RNase H and DNase I domains were also searched using the same parameters.

The databases Interpro, Pfam, SMART, Superfamily, and PROSITE were used to obtain PDB codes from KH domain families and families related to KH domains. Redundant codes were removed. Non-redundant codes were inspected for binding to RNA and binding to RNA in the α' loop. Alignment of selected KH domains (Figure 38) was performed by MatchMaker in Chimera (Pettersen et al, 2004).

6 SUMMARY AND PERSPECTIVES

Post-transcriptional regulation is an important aspect of gene expression control in the life of a cell. In *E. coli*, the RNA degradosome serves as an integral post-transcriptional regulator by influencing the lifetime of mRNAs (Cohen & McDowall, 1997) as well as facilitating small RNA mediated control of targeted transcripts. Aside from unstable RNAs such as sRNAs and mRNAs, stable RNAs — such as rRNAs or tRNAs — may also be degraded as a response to misfolding or stress by RNase E within the degradosome (Sulthana et al, 2016). The RNA degradosome can therefore be viewed as a type of surveillance machinery. Though much research has been performed on the RNA degradosome since its discovery, some finer details of its mechanisms remain to be addressed. It is known that in *E. coli*, the RNA degradosome is a large multienzyme complex with a canonical core consisting of RNase E, RhlB, enolase, and PNPase; other proteins can also associate and sometimes dissociate depending on growth conditions (Prud'homme-Généreux et al, 2004). The hydrolytic RNA activity of the machine is provided by RNase E, which can be divided into two domains, the N-terminal domain and the C-terminal domain. The N-terminal domain acts as an endoribonuclease which cleaves RNA. Cleavage of the targeted RNA may induce subsequent rapid degradation or create a matured form of structured RNAs. The C-terminal domain, on the other hand, is an unstructured scaffold upon which the other components of the degradosome bind. There are two pathways by which the RNase E catalytic N-terminal domain recognizes its substrate, the 5' sensing pathway and the direct entry pathway. The 5' sensing pathway recognizes a 5'-monophosphate either in cis or trans. Recognition of the 5'-monophosphate leads to domain closure and formation of an RNA binding channel that presents single-stranded

RNA to the catalytic site, leading to cleavage of the substrate (Figure 5). In the direct entry pathway, structured RNA may associate with RNase E directly in a way which promotes cleavage of the RNA, bypassing the 5' sensor. Though new structural information reveals how RNase E may bind to target RNA in the direct entry pathway (Bandyra et al, 2018), exactly how RNA cleavage is induced by this activity mode remains to be visualised.

The C-terminal of RNase E is mostly unstructured save for a few microdomains. These elements include the membrane targeting sequence (MTS), arginine rich RNA binding site 1 (AR1), RhlB binding site (RBS), arginine rich RNA binding site 2 (AR2), enolase binding site (EBS), and PNPase binding site (PBS) (Figure 4). Each microdomain has a unique function, but the details regarding interactions between individual units of the degradosome remain unknown. After endonucleolytic cleavage of an RNA by the N-terminal domain of RNase E, the RNA is then passed off to PNPase at the C-terminus for processive degradation in the 3' to 5' direction (Bandyra & Luisi, 2018). As PNPase requires single stranded RNA for processivity, the DEAD box helicase RhlB facilitates its activity by unwinding complex structures which are likely to exist in intergenic regions (Khemici & Carpousis, 2004). The MTS serves to attach the degradosome machinery to the inner membrane, potentially resulting in a temporal lag between the transcription of mRNAs and their degradation (Mackie 2013). Enolase's function in the degradosome has been largely speculative, though evolutionary comparisons of all degradosomes has shown that glycolytic enzymes or TCA cycle enzymes are often components of the assembly. These associations suggest some link between cellular energy status and RNA degradation (Chandran & Luisi, 2006; Hardwick et al, 2011; Kovacs et al, 2005). More recently, enolase binding to the degradosome was found to be important for the activity of certain sRNAs (Morita et al, 2004) and for control of cell division (Murashko and Lin-Chao, 2017). Though

the functions of individual components of the degradosome are well described, the information surrounding their cooperation and communication has largely been speculative.

One key protein involved in post-transcriptional regulation that can work in conjunction with the degradosome is Hfq. Hfq is an RNA chaperone responsible for protecting small regulatory RNAs (sRNAs) from degradation and facilitating sRNA pairing to target mRNAs (Vogel & Luisi, 2011). Hfq has been shown to associate with the RNA degradosome both in the N-terminal and the C-terminal domains (Bandyra et al, 2013, Ikeda et al, 2011; Worrall et al., 2008). Hfq has three modes of binding to RNA, involving the proximal and distal faces and the circumferential rim (Figure 3). The proximal face prefers to bind poly-U RNA while the distal face prefers to bind AAU repeats (Vogel & Luisi, 2011). Incidentally, poly-U RNA appears more often in sRNAs as a Rho-independent transcription terminator while AAU repeats occur more often in mRNAs. Hence, Hfq may bind sRNAs on one face while its target mRNA binds on the other face, bringing the two RNAs in proximity with each other. Aside from binding to an sRNA on one face and mRNA on the other face, Hfq has been shown to bind to some sRNAs with both faces (Schu et al, 2015). These small RNAs have, in general, a smaller pool of targets and are responsible for more specific cellular responses. Though Hfq associations with these RNAs are well described, there are few experimental structures of Hfq bound to RNAs. Currently, there is only one structure of Hfq bound to an sRNA, RydC, alongside some structures of Hfq bound to synthetic RNA. RydC happens to be an sRNA which only binds the proximal face of Hfq, and thus, the structure and therefore the details of an sRNA which associates with both the proximal and distal faces of Hfq remain undescribed.

This thesis aimed to answer some of the questions regarding post-transcriptional regulation which remain unresolved (Figure 42). One of the questions I wanted to address was

how RhlB inactivation affects the function of the degradosome, both in the regulatory mode and in the degradative function. I hoped that by understanding the specific role of RhlB in more depth, it would be possible to come to a better understanding of the overall communication between the individual degradosome components. The overall architecture of the RNase E C-terminal domain attached to its individual degradosome components has been the subject of previous investigation (Bruce et al, 2018). With the advent of the cryo-EM resolution revolution, I hoped to further understand the molecular interactions within the individual components of the degradosome. In addition, I hoped to illuminate the mechanism by which an sRNA can bind to both faces of Hfq. As Hfq has been shown to bind to the C-terminal domain of the degradosome (Ikeda et al, 2011), most likely through shared binding of RNA (Worrall et al., 2008), I also wanted to see how Hfq would associate with a segment of the degradosome. In the N-terminal domain, the small domain of RNase E has been shown to interact with duplex RNA, and one question is whether this could be a shared mode of interaction with a stem-loop structure exposed on the surface of Hfq to form a transient hand-over complex.

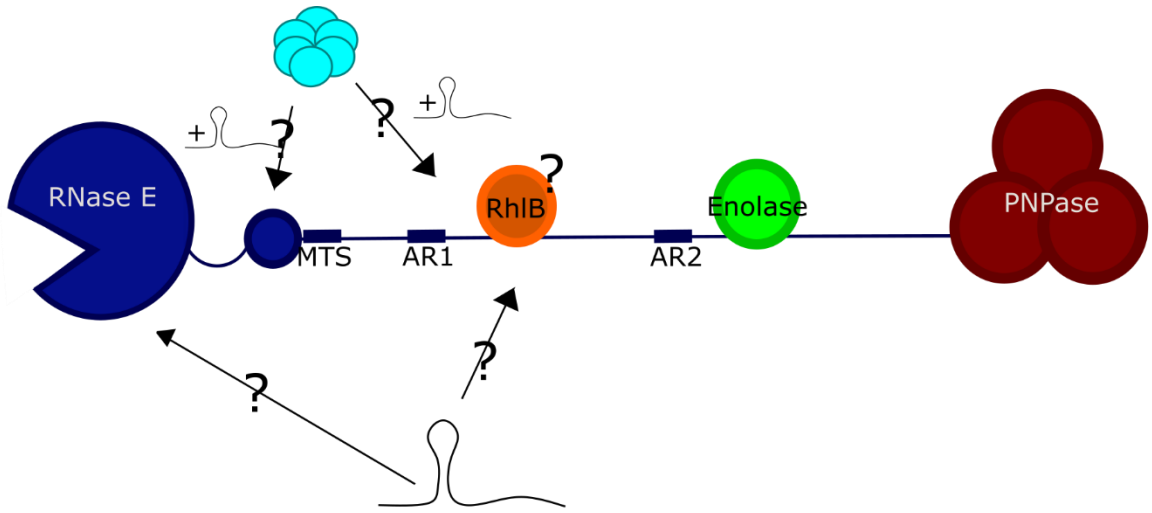


Figure 42: Unanswered questions investigated in this thesis. The primary purpose of this thesis was to examine the structure and function of RhlB, as well as its method of interactions with the degradosome and RNAs. Alongside RhlB, Hfq's association with the degradosome C-terminal as well as the class II sRNA ChiX was investigated. The direct entry cleavage pathway of RNase E was also briefly explored.

To answer these questions, I employed a point mutation in the DEAD box of RhlB, mutating the glutamate to glutamine (E166Q). This point mutation was previously shown in RhlB homologues to create an RNA clamp which is defective in the release of unwound RNA (Xiol et al, 2014). This chromosomal point mutation was generated with the help of our collaborators, and I sought to characterize whether this mutation affects the growth of mutant cells. In my experiments, I found that the mutant strains seemingly grew faster than the WT strains in rich media, except when subjected to iron stress. Upon further speculation, the genomic background of WT cells may have the greatest contribution to the difference in growth rate. Point mutations in the DEAD box of RhlB may be insufficient to cause a disturbance in

the cell large enough to alter the growth rate. Thus, further investigations into the biochemical mechanism of the RhlB DQAD mutations will be necessary for its characterization.

To that point, I decided to create the same mutant in a plasmid used for overexpression and purification. As RhlB's ATPase rate changes when bound to RNase E, unbound RhlB may not be representative of how RhlB behaves in the context of the degradosome. Therefore, the mutant RhlB was made in a plasmid which co-expressed RNase E 603-850. The *E. coli* expression host was co-transformed with this plasmid and another plasmid expressing enolase. Enolase was also included to form the degradosome sub-assembly for characterisation, as that glycolytic enzyme has been speculated to change the AR2 RNA binding site and therefore change the affinity of the RNase E C-terminal to RNA. Thus, all three components were included to mimic the RNA binding capabilities of the degradosome C-terminal domain. The three-component RhlB, RNase E 603-850, and enolase assembly was named the ternary complex for convenience. Mutant ternary complex containing RhlB E166Q was shown to bind RNA 10 x more strongly than WT RhlB. Following this observation, I decided to use this mutation to find transient RhlB-RNA interactions in the cell. Mutants with affinity tagged RhlB E166Q were kindly provided by our collaborators. I used UV irradiation to crosslink proteins and nucleotides together and used denaturing buffer to separate RhlB from the other degradosome components. Initial crosslinking and purification seemed successful, as RhlB was shown to be associated with RNA which degraded during alkaline hydrolysis. We hope to further optimize this protocol and sequence the RNA that was bound to RhlB to identify novel substrates that the helicase acts upon in the degradosome.

For a deeper understanding of the interactions of the degradosome, several new experiments can be performed in the future. To assess the population of RNA which is bound

to RhlB, crosslinking and deep sequencing may be performed. For the deep sequencing experiments, cell lines that are wild type or contain mutant RhlB may be subjected to different stress conditions (such as acid stress, iron stress, sugar stress, etc) during mid log and stationary growth phases. Total RNA may then be harvested after 20 minutes of stress and sequenced. Rates of RNA degradation can be evaluated by quantify RNA species after stopping transcription with rifampicin. The amount of RNA in mutant cell lines can be compared to WT cell lines to understand how inactivation of RhlB affects RNA population distribution and RNA degradation rates. For the crosslinking experiment, the initial steps performed will be the same as the crosslinking steps outlined in Chapter 3. Following pulldown purifications under denaturing conditions, RhlB bound to RNA should then undergo sequencing to identify which RNAs may associate with RhlB during its lifetime (Figure 43). These two experiments together will yield a more complete picture of RhlB function in the degradosome.

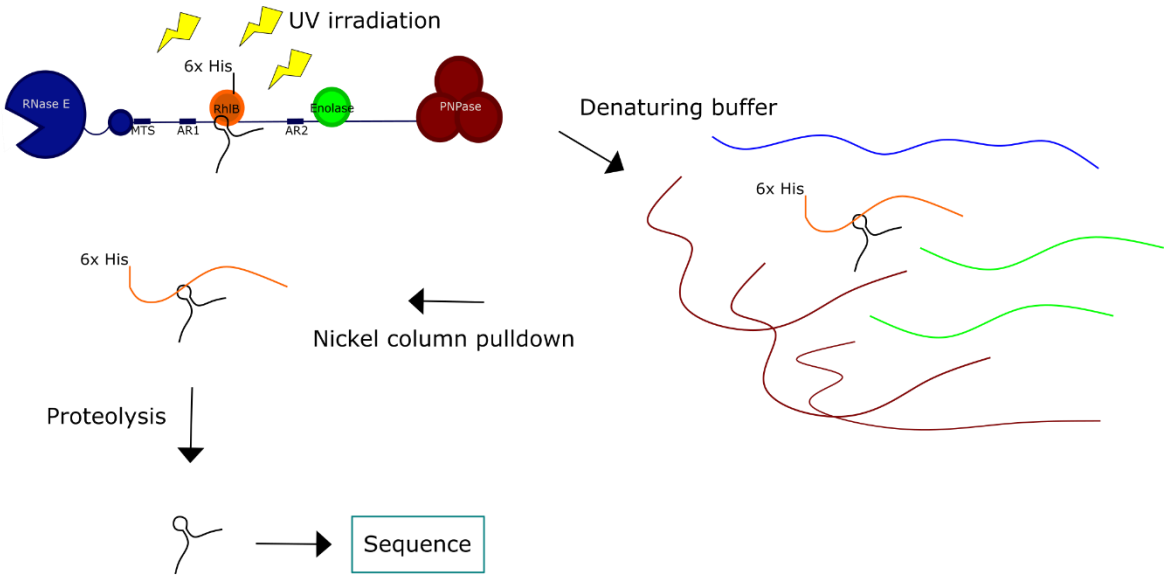


Figure 43: Experimental outline of UV crosslinking and pulldown. Cells containing affinity tagged mutant RhlB are irradiated with UVC, then lysed using a denaturing buffer. RhlB crosslinked to RNA is then separated from other proteins with affinity purification. Following purification, RhlB is then lysed and the bound RNA may then be sequenced and analyzed.

In an attempt to understand the structure of the C-terminal of the degradosome, I chose to utilize the technique cryo-EM as it can image flexible and conformationally heterogeneous proteins. As larger proteins often provide better signal and therefore higher resolution reconstructions, I decided to add Hfq and ChiX — an sRNA which mediates chitin metabolism that binds to both sides of Hfq — to the ternary complex to make a supercomplex. As Hfq has been shown to bind the unstructured region of the degradosome through RNA (Worrall et al., 2008), we hoped that with one complex we could resolve multiple unknowns, namely the organization and communication of the degradosome C-terminus, the association of Hfq to the degradosome C-terminus, and how ChiX may bind to both sides of Hfq. During imaging, I

observed that the supercomplex was prone to disassociation and aggregation on the carbon support of the EM grid. To remedy these artefacts, the supercomplex was reinforced with glutaraldehyde crosslinking and treated proteins were frozen on gold grids without carbon support. After image collection, I had obtained a map of a subcomplex that may be Hfq bound to ChiX. Using the map, I attempted to fit the two molecules to estimate how ChiX may bind to Hfq. One characteristic of the map was that the majority of ChiX seems to be located on one surface of the Hfq, which is most likely the distal face. This may result in protecting the sRNA from non-specific binding, or to enable the seed region to contact more RNA to kinetically facilitate target recognition and response. Though the general organization of the ChiX bound to Hfq may be speculated upon, a conclusive description of the interactions of ChiX bound to Hfq requires a higher resolution structure.

Future experiments may be conducted to further understand the structure of the ternary complex as well as the structure of ChiX bound to Hfq. There remains unprocessed cryo-EM data containing a complex which is of comparable size to the ternary complex. These images should be processed for a 3D model of what the ternary complex may look like. In addition to data processing, other methods of sample preparation may be performed. The GraFix gradient fixation negative staining approach may be performed to purify and crosslink the supercomplex together (Kastner et al, 2007). Other grid surface materials, such as graphene oxide and PEGylated gold, may be used to prevent the complex from dissociating and precipitating on carbon supports (Meyerson et al, 2014; Wilson et al, 2009). In addition, crystallography may be employed as a technique to visualize ChiX binding to Hfq.

One aspect of the degradosome which does not involve the C-terminal unstructured region is the direct entry of certain RNAs into the cleavage site of RNase E. These RNAs do

not require a 5' monophosphate to activate the endonucleolytic activity of the RNase E N-terminus. Recently, Bandyra et al published a structure of RNase E N-terminal domain bound to RprA which showed a novel RNA binding site on the RNase E small domain and DNase I domain. Concurrently, another paper showed that the RNase E small domain is another member of the KH domain of RNA binding proteins (Pereira & Lupas, 2018). To further examine the link between RNase E small domains and the KH domain, I carried out a bioinformatic analysis which included sequential and structural analysis. I utilized jackhmmer to iteratively search the UniProtKB database for sequences which may be distantly related to the query (RNase E 415-531). I also gathered all the PDB codes of KH domains in the Interpro, Pfam, SMART, Superfamily, and PROSITE databases to search for structural similarities in RNA binding between the new structure of RNase E bound to RprA and other previously known KH domains. Through this bioinformatic analysis, a similarity between the RNase E small domain and previously described KH domains were found both in sequence and in structure. In the sequences found, only 209 out of 33,000+ sequences were not related to the KH domain. The KH motif of IGxGxxIK/R was prominent in the HMMER logo generated from all sequences. Structurally, the RNase E small domain bound to RprA had a similar binding geometry to 412 out of 677 structures of KH domains bound to RNA. These findings support the previously reported conclusions that the RNase E small domain is a member of the KH domain family, as well as the potential of the RNase E small domain to act as an RNA sensor in the direct entry pathway for RNA degradation. The mode of interaction of the small domain and RNase H domains of RNase E with duplex RNA might occur in effector complexes. For example, it can be envisaged that duplex regions of sRNA associated with Hfq could be exposed to engage RNase E, while the seed region could help to present the mRNA target for cleavage.

To further understand the complex system that is the degradosome, additional studies can be envisaged suggested from some of the studies presented here. Structural information about the degradosome, its subassemblies and cognate complexes with target RNAs can offer insights into the dynamics of such a large and mobile molecular machine. The degradosome's wide ranging activity as well as speed of action and precision in gene expression control make it an extraordinarily intriguing target for investigation. I hope that through the scrutiny of the RNA degradosome, both in vivo and in vitro with advanced structural approaches and integrative methods, a more complete picture of the prokaryotic post-transcriptional regulatory landscape can be revealed.

7 REFERENCES

Aguirre AA, Vicente AM, Hardwick SW, Alvelos DM, Mazzon RR, Luisi BF & Marques M V. (2017) Association of the cold shock DEAD-box RNA helicase RhlE to the RNA degradosome in Caulobacter crescentus. *J. Bacteriol.* **199:** e00135-17

Aiba H, Matsuyama SI, Mizuno T & Mizushima S (1987) Function of micF as an antisense RNA in osmoregulatory expression of the ompF gene in Escherichia coli. *J. Bacteriol.* **169:** 3007–3012

Aït-Bara S & Carpousis AJ (2015) RNA degradosomes in bacteria and chloroplasts: Classification, distribution and evolution of RNase E homologs. *Mol. Microbiol.* **97:** 1021–1035

Aït-Bara S, Carpousis AJ & Quentin Y (2015) RNase E in the γ-Proteobacteria: conservation of intrinsically disordered noncatalytic region and molecular evolution of microdomains. *Mol. Genet. Genomics* **290:** 847–862

Andersen J & Delihas N (1990) micF RNA Binds to the 5′ End of ompF mRNA and to a Protein from Escherichia coli. *Biochemistry* **29:** 9249–9256

Andersen J, Delihas N, Ikenaka K, Green PJ, Pines O, Ilercil O & Inouye M (1987) The isolation and characterization of RNA coded by the micF gene in Escherichia coli. *Nucleic Acids Res.* **15:** 2089–2101

Antal M, Bordeau V, Douchin V & Felden B (2005) A small bacterial RNA regulates a putative ABC transporter. *J. Biol. Chem.* **280:** 7901–7908

Bai X-C, Fernandez IS, McMullan G & Scheres SHW (2013) Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *eLife* **2:** e00461

Antal M, Bordeau V, Douchin V & Felden B (2005) A small bacterial RNA regulates a putative ABC transporter. *J. Biol. Chem.* **280:** 7901–7908

Bandyra KJ, Bouvier M, Carpousis AJ & Luisi BF (2013) The social fabric of the RNA degradosome. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1829:** 514–522

Bandyra KJ & Luisi BF (2018) RNase E and the High-Fidelity Orchestration of RNA Metabolism. *Regul. with RNA Bact. Archaea***:** 3–18

Bandyra KJ, Wandzik JM & Luisi BF (2018) Substrate Recognition and Autoinhibition in the Central Ribonuclease RNase E. *Mol. Cell* **72:** 275

Bandyra KJ, Sinha D, Syrjanen J, Luisi BF & De Lay NR (2016) The ribonuclease polynucleotide phosphorylase can interact with small regulatory RNAs in both protective and degradative modes. *RNA* **22:** 360–72

Bandyra KJ, Said N, Pfeiffer V, Górna MWW, Vogel J & Luisi BFF (2012) The Seed Region of a Small RNA Drives the Controlled Destruction of the Target mRNA by the Endoribonuclease RNase E. *Mol. Cell* **47:** 943–953

Bass RB, Strop P, Barclay M & Rees DC (2002) Crystal structure of Escherichia coli MscS, a voltage-modulated and mechanosensitive channel. *Science* **298:** 1582–1587

Battesti A, Majdalani N & Gottesman S (2011) The RpoS-Mediated General Stress Response in Escherichia coli . *Annu. Rev. Microbiol.* **65:** 189–213

Bedouelle H, Schmeissner U, Hofnung M & Rosenberg M (1982) Promoters of the malEFG and malK-lamB operons in *Escherichia coli* K12. *J. Mol. Biol.* **161:** 519–531

Beich-Frandsen M, Večerek B, Konarev P V., Sjöblom B, Kloiber K, Hämmerle H, Rajkowitsch L, Miles AJ, Kontaxis G, Wallace BA, Svergun DI, Konrat R, Bläsi U & Djinović-Carugo K (2011) Structural insights into the dynamics and function of the C-terminus of the E. coli RNA chaperone Hfq. *Nucleic Acids Res.* **39:** 4900–4915

Beisel CL & Storz G (2011) The Base-Pairing RNA Spot 42 Participates in a Multioutput Feedforward Loop to Help Enact Catabolite Repression in *Escherichia coli*. *Mol. Cell* **41:** 286–297

Betzig E, Patterson G, Sougrat R, Lindwasser O, Olenych S, Bonifacino J, Davidson M, Lippincott-Schwartz J & Hess H (2006) Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. *Science* **313:** 1642-1645

Biesiada M, Purzycka KJ, Szachniuk M, Blazewicz J & Adamiak RW (2016) Automated RNA 3D Structure Prediction with RNAComposer. *Methods Mol. Biol.* **1490:** 199–215.

Bobrovskyy M & Vanderpool CK (2016) Diverse mechanisms of post-transcriptional repression by the small RNA regulator of glucose-phosphate stress. *Mol. Microbiol.* **99:** 254–73

Bobrovskyy M & Vanderpool CK (2014) The small RNA SgrS: roles in metabolism and pathogenesis of enteric bacteria. *Front. Cell. Infect. Microbiol.* **4:** 61

Bordeau V & Felden B (2014) Curli synthesis and biofilm formation in enteric bacteria are controlled by a dynamic small RNA module made up of a pseudoknot assisted by an RNA chaperone. *Nucleic Acids Res.* **42:** 4682–4696

Bouvier M & Carpousis AJ (2011) A tale of two mRNA degradation pathways mediated by RNase E. *Mol. Microbiol.* **82:** 1305–1310

Bruce HA, Du D, Matak-Vinkovic D, Bandyra KJ, Broadhurst RW, Martin E, Sobott F, Shkumatov A V. & Luisi BF (2018) Analysis of the natively unstructured RNA/protein-recognition core in the *Escherichia coli* RNA degradosome and its interactions with regulatory RNA/Hfq complexes. *Nucleic Acids Res.* **46:** 387–402

Callaghan AJ, Aurikko JP, Ilag LL, Günter Grossmann J, Chandran V, Kühnel K, Poljak L, Carpousis AJ, Robinson C V., Symmons MF & Luisi BF (2004) Studies of the RNA

degradosome-organizing domain of the Escherichia coli ribonuclease RNase E. *J. Mol. Biol.* **340:** 965–979

Callaghan AJ, Marcaida MJ, Stead JA, McDowall KJ, Scott WG & Luisi BF (2005) Structure of Escherichia coli RNase E catalytic domain and implications for RNA turnover. *Nature* **437:** 1187–1191

Cameron TA & De Lay NR (2016) The phosphorolytic exoribonucleases polynucleotide phosphorylase and RNase PH stabilize sRNAs and facilitate regulation of their mRNA targets. *J. Bacteriol.* **198:** 3309–3317

Carpousis AJ (2007) The RNA Degradosome of Escherichia coli : An mRNA-Degrading Machine Assembled on RNase E . *Annu. Rev. Microbiol.* **61:** 71–87

Chandran V & Luisi BF (2006) Recognition of enolase in the Escherichia coli RNA degradosome. *J. Mol. Biol.* **358:** 8–15

Chandran V, Poljak L, Vanzo NF, Leroy A, Miguel RN, Fernandez-Recio J, Parkinson J, Burns C, Carpousis AJ & Luisi BF (2007) Recognition and Cooperation Between the ATP-dependent RNA Helicase RhlB and Ribonuclease RNase E. *J. Mol. Biol.* **367:** 113–132

Chao Y, Papenfort K, Reinhardt R, Sharma CM & Vogel J (2012) An atlas of Hfq-bound transcripts reveals 3′ UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J.* **31:** 4005–4019

Chao Y & Vogel J (2016) A 3’ UTR-Derived Small RNA Provides the Regulatory Noncoding Arm of the Inner Membrane Stress Response. *Mol. Cell* **61:** 352–363

Chao Y & Vogel J (2010) The role of Hfq in bacterial pathogens. *Curr. Opin. Microbiol.* **13:** 24–33

Chareyre S & Mandin P (2018) Bacterial Iron Homeostasis Regulation by sRNAs. *Microbiol. Spectr.* **6:** 267–281

Cherepanov PP & Wackernagel W (1995) Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene* **158:** 9–14

Chowdhury A, Raju KK, Kalurupalle S & Tharun S (2012) Both Sm-domain and C-terminal extension of Lsm1 are important for the RNA-binding activity of the Lsm1-7-Pat1 complex. *RNA* **18:** 936–944

Chung HJ, Bang W & Drake MA (2006) Stress response of Escherichia coli. *Compr. Rev. Food Sci. Food Saf.* **5:** 52–64

Clarke JE, Kime L, Romero A. D & McDowall KJ (2014) Direct entry by RNase e is a major pathway for the degradation and processing of RNA in Escherichia coli. *Nucleic Acids Res.* **42:** 11733–11751

Cohen SN & McDowall KJ (1997) RNase E: Still a wonderfully mysterious enzyme. *Mol. Microbiol.* **23:** 1099–1106

Coleman J, Green PJ & Inouye M (1984) The use of RNAs complementary to specific mRNAs to regulate the expression of individual bacterial genes. *Cell* **37:** 429–436

Cordin O, Banroques J, Tanner NK & Linder P (2006) The DEAD-box protein family of RNA helicases. *Gene* **367:** 17–37

Copper-tem-grid-500x500 (2019) Available at: https://5.imimg.com/data5/OC/WF/AV/SELLER-6427745/copper-tem-grid-500x500.png [Accessed December 16, 2019]

Datsenko KA & Wanner BL (2000) One-Step Inactivation of Chromosomal Genes in *Escherichia Coli* K-12 Using PCR Products. *Proc. Nat. Acad. Sci.* 97, **12:** 6640–6645.

De Lay N, Schu DJ & Gottesman S (2013) Bacterial Small RNA-Based Negative Regulation: Hfq and Its Accomplices. *J. Biol. Chem.* **288:** 12.

Del Favero M, Mazzantini E, Briani F, Zangrossi S, Tortora P & Dehò G (2008) Regulation of Escherichia coli polynucleotide phosphorylase by ATP. *J. Biol. Chem.* **283:** 27355–27359

Delihas N & Forst S (2001) MicF: An antisense RNA gene involved in response of Escherichia coli to global stress factors. *J. Mol. Biol.* **313:** 1–12

Deng Z, Liu Z, Bi Y, Wang X, Zhou D, Yang R & Han Y (2014) Rapid degradation of Hfq-free RyhB in Yersinia pestis by PNPase independent of putative ribonucleolytic complexes. *BioMed Res. Int.* **2014:** 798918

Desnoyers G & Masse E (2012) Noncanonical repression of translation initiation through small RNA recruitment of the RNA chaperone Hfq. *Genes Dev.* **26:** 726–739

Desnoyers G, Morissette A, Prévost K & Massé E (2009) Small RNA-induced differential degradation of the polycistronic mRNA iscRSUA. *EMBO J.* **28:** 1551–61

Diestra E, Cayrol B, Arluison V & Risco C (2009) Cellular Electron Microscopy Imaging Reveals the Localization of the Hfq Protein Close to the Bacterial Membrane. *PLOS ONE* **4:** e8301

Dimastrogiovanni D, Fröhlich KS, Bandyra KJ, Bruce HA, Hohensee S, Vogel J & Luisi BF (2014) Recognition of the small regulatory RNA RydC by the bacterial Hfq protein. *eLIFE* **3**: e05375

Domínguez-Malfavón L, Islas LD, Luisi BF, García-Villegas R & García-Mena J (2013) The assembly and distribution in vivo of the Escherichia coli RNA degradosome. *Biochimie* **95:** 2034–2041

Emsley P, Lohkamp B, Scott WG & Cowtan K (2010) Features and development of *Coot*. *Acta Crystallogr. D* **66:** 486–501

Faner MA & Feig AL (2013) Identifying and characterizing Hfq-RNA interactions. *Methods* **63:** 144–159

Fender A, Elf J, Hampel K, Zimmermann B & Wagner EGH (2010) RNAs actively cycle on the Sm-like protein Hfq. *Genes Dev.* **24:** 2621–6

Fernandez-Leiro R & Scheres SHW (2017) A pipeline approach to single-particle processing in RELION. *Acta Crystallogr. D* **73:** 496–502

Figueroa-Bossi N, Valentini M, Malleret L, Fiorini F & Bossi L (2009) Caught at its own game: regulatory small RNA inactivated by an inducible transcript mimicking its target. *Genes Dev.* **23:** 2004–15

Fröhlich KS, Papenfort K, Berger AA & Vogel J (2012) A conserved RpoS-dependent small RNA controls the synthesis of major porin OmpD. *Nucleic Acids Res.* **40:** 3623–3640

Fröhlich KS, Papenfort K, Fekete A & Vogel J (2013) A small RNA activates CFA synthase by isoform-specific mRNA stabilization. *EMBO J.* **32:** 2963–2979

Fröhlich KS & Vogel J (2009) Activation of gene expression by small RNA. *Curr. Opin. Microbiol.* **12:** 674–682

G3727_cryo-grid-boxes-v2_1200px (2019) Available at: http://www.agarscientific.com/media/catalog/product/cache/1/image/17f82f742ffe127f42dca9de82fb58b1/g/3/g3727_cryo-grid-boxes-v2_1200px.jpg [Accessed December 16, 2019]

Geissmann TA & Touati D (2004) Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator. *EMBO J.* **23:** 396–405

Hadjeras L, Poljak L, Bouvier M, Morin-Ogier Q, Canal I, Cocaign-Bousquet M, Girbal L & Carpousis AJ (2019) Detachment of the RNA degradosome from the inner membrane of Escherichia coli results in a global slowdown of mRNA degradation, proteolysis of RNase E and increased turnover of ribosome-free transcripts. *Mol. Microbiol.* **111:** 1715–1731

Hardwick SW, Chan VSY, Broadhurst RW & Luisi BF (2011) An RNA degradosome assembly in Caulobacter crescentus. *Nucleic Acids Res.* **39:** 1449–1459

Hell S & Wichmann J (1994) Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics Letters* **19:** 780-782

Henderson CA, Vincent HA, Casamento A, Stone CM, Phillips JO, Cary PD, Sobott F, Gowers DM, Taylor JE & Callaghan AJ (2013) Hfq binding changes the structure of Escherichia coli small noncoding RNAs OxyS and RprA, which are involved in the riboregulation of rpoS. *RNA* **19:** 1089–1104

Hermann H, Fabrizio P, Raker VA, Foulaki K, Hornig H, Brahms H & Lührmann R (1995) snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein-protein interactions. *EMBO J.* **14:** 2076–88

Hoekzema M, Romilly C, Holmqvist E & Wagner EGH (2019) Hfq-dependent mRNA unfolding promotes sRNA -based inhibition of translation . *EMBO J.* **38:** e101199

HOME-DUAL__40784.1549918163 (2019) Available at: https://cdn11.bigcommerce.com/s-ufhcuzfxw9/images/stencil/1280x1280/products/14274/17346/HOME-DUAL__40784.1549918163.jpg?c=2&imbypass=on&imbypass=on [Accessed December 16, 2019]

Houdellier F, de Knoop L, Gatel C, Masseboeuf A, Mamishin S, Taniguchi Y, Delmas M, Monthioux M, Hÿtch M & Snoeck E (2015) Development of TEM and SEM high

brightness electron guns using cold-field emission from a carbon nanotip. *Ultramicroscopy* **151:** 107-115

Hunter CA & Anderson HL (2009) What is Cooperativity? *Angew. Chem.* **48:** 7488–7499

Ikeda Y, Yagi M, Morita T & Aiba H (2011) Hfq binding at RhlB-recognition region of RNase E is crucial for the rapid degradation of target mRNAs mediated by sRNAs in Escherichia coli. *Mol. Microbiol.* **79:** 419–432

Iqbal M & Fazal-e-Aleem (2005) Theory And Design Of Thermionic Electron Beam Guns. *AIP Conference Proceedings* **748:** 376

Ishikawa H, Otaka H, Maki K, Morita T & Aiba H (2012) The functional Hfq-binding module of bacterial sRNAs consists of a double or single hairpin preceded by a U-rich sequence and followed by a 3' poly(U) tail. *RNA* **18:** 1062–1074

Jager S (2001) An mRNA degrading complex in Rhodobacter capsulatus. *Nucleic Acids Res.* **29:** 4581–4588

Jarmoskaite I & Russell R (2011) DEAD-box proteins as RNA helicases and chaperones. *Wiley Interdiscip. Rev. RNA* **2:** 135–152

Jia M, Horita S, Nagata K & Tanokura M (2010) An Archaeal Dim2-Like Protein, aDim2p, Forms a Ternary Complex with a/eIF2α and the 3′ End Fragment of 16S rRNA. *J. Mol. Biol.* **398:** 774-785

Jiang X & Belasco JG (2004) Catalytic activation of multimeric RNase E and RNase G by 5′-monophosphorylated RNA. *Proc. Natl. Acad. Sci. U. S. A.* **101:** 9211–9216

Jonas S & Izaurralde E (2013) The role of disordered protein regions in the assembly of decapping complexes and RNP granules. *Genes Dev.* **27:** 2628–2641

Joyce SA & Dreyfus M (1998) In the absence of translation, RNase E can bypass 5' mRNA stabilizers in Escherichia coli. *J. Mol. Biol.* **282:** 241–254

Kambach C, Walke S, Young R, Avis JM, de la Fortelle E, Raker VA, Lührmann R, Li J & Nagai K (1999) Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell* **96:** 375–87

Kastner B, Fischer N, Golas M, Sander B, Dube P, Boehringer D, Hartmuth K, Deckert J, Hauer F, Wolf E, Uchtenhagen H, Urlaub H, Herzog F, Peters J, Poerschke D, Lührmann R & Stark H (2007) GraFix: sample preparation for single-particle electron cryomicroscopy. *Nature Methods* **5:** 53-55

Kavita K, de Mets F & Gottesman S (2018) New aspects of RNA-based regulation by Hfq and its partner sRNAs. *Curr. Opin. Microbiol.* **42:** 53–61

Kawamoto H, Koide Y, Morita T & Aiba H (2006) Base-pairing requirement for RNA silencing by a bacterial small RNA and acceleration of duplex formation by Hfq. *Mol. Microbiol.* **61:** 1013–1022

Kessler JR, Cobe BL & Richards GR (2017) Stringent Response Regulators Contribute to Recovery from Glucose Phosphate Stress in Escherichia coli. *Appl. Environ. Microbiol.* **83:** e01636–17

Khemici V & Carpousis AJ (2003) The RNA degradosome and poly(A) polymerase of Escherichia coli are required in vivo for the degradation of small mRNA decay intermediates containing REP-stabilizers. *Mol. Microbiol.* **51:** 777–790

Khemici V, Poljak L, Luisi BF & Carpousis AJ (2008) The RNase E of *Escherichia coli* is a membrane-binding protein. *Mol. Microbiol.* **70:** 799–813

Knapek E & Dubochet J (1980) Beam damage to organic material is considerably reduced in cryo-electron microscopy. *Journal of Molecular Biology* **141:** 147-161

Koslover DJ, Callaghan AJ, Marcaida MJ, Garman EF, Martick M, Scott WG & Luisi BF (2008) The Crystal Structure of the Escherichia coli RNase E Apoprotein and a Mechanism for RNA Degradation. *Structure* **16:** 1238–1244

Kovacs L, Csanadi A, Megyeri K, Kaberdin VR & Miczak A (2005) Mycobacterial RNase E-associated proteins. *Microbiol. Immunol.* **49:** 1003–1007

Kruger D, Schneck P & Gelderblom H (2000) Helmut Ruska and the visualisation of viruses. *The Lancet* **355:** 1713-1717

Kucera NJ, Hodsdon ME & Wolin SL (2011) An intrinsically disordered C terminus allows the La protein to assist the biogenesis of diverse noncoding RNA precursors. *Prot. Nat. Acad. Sci.* **108:** 1308–1313

Lalaouna D, Carrier M-C, Semsey S, Brouard J-S, Wang J, Wade JT & Massé E (2015) A 3' external transcribed spacer in a tRNA transcript acts as a sponge for small RNAs to prevent transcriptional noise. *Mol. Cell* **58:** 393–405

Lalaouna D, Eyraud A, Devinck A, Prévost K & Massé E (2019) GcvB small RNA uses two distinct seed regions to regulate an extensive targetome. *Mol. Microbiol.* **111:** 473-486

Laskowski RA, MacArthur MW, Moss DS & Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr* **26:** 283–291

Lee K & Cohen SN (2003) A Streptomyces coelicolor functional orthologue of Escherichia coli RNase E shows shuffling of catalytic and PNPase-binding domains. *Mol. Microbiol.* **48:** 349–360

Lehnik-Habrink M, Newman J, Rothe FM, Solovyova AS, Rodrigues C, Herzberg C, Commichau FM, Lewis RJ & Stülke J (2011) RNase Y in Bacillus subtilis: A natively disordered protein that is the functional equivalent of RNase E from Escherichia coli. *J. Bacteriol.* **193:** 5431–5441

Leung AKW, Nagai K & Li J (2011) Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis. *Nature* **473:** 536–539

Lewis HA, Chen H, Edo C, Buckanovich RJ, Yang YY, Musunuru K, Zhong R, Darnell RB & Burley SK (1999) Crystal structures of Nova-1 and Nova-2 K-homology RNA-binding domains. *Structure* **7:** 191–203

Linder P & Jankowsky E (2011) From unwinding to clamping - the DEAD box RNA helicase family. *Nat. Rev. Mol. Cell Biol.* **12:** 505–516

Linder P, Lasko PF, Ashburner M, Leroy P, Nielsen PJ, Nishi K, Schnier J & Slonimski PP (1989) Birth of the D-E-A-D box. *Nature* **337:** 121–122

Link TM, Valentin-Hansen P & Brennana RG (2009) Structure of Escherichia coli Hfq bound to polyriboadenylate RNA. *Proc. Natl. Acad. Sci. U. S. A.* **106:** 19292–19297

Lloyd CR, Park S, Fei J & Vanderpool CK (2017) The small protein SgrT controls transport activity of the glucose-specific phosphotransferase system. *J. Bacteriol.* **199:** e00869-16

Mackie GA (2013) RNase E: At the interface of bacterial RNA processing and decay. *Nat. Rev. Microbiol.* **11:** 45–57

Marbach A & Bettenbrock K (2012) lac operon induction in Escherichia coli: Systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA. *Journal of Biotechnology* **157**: 82-88

Massé E, Escorcia FE & Gottesman S (2003) Coupled degradation of a small regulatory RNA and its mRNA targets in Escherichia coli. *Genes Dev.* **17:** 2374–83

Massé E & Gottesman S (2002) A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli. *Prot. Nat. Acad. Sci.* **99:** 4620–5

Massé E, Vanderpool CK & Gottesman S (2005) Effect of RyhB small RNA on global iron use in Escherichia coli. *J. Bacteriol.* **187:** 6962–6971

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC & Read RJ (2007) *Phaser* crystallographic software. *J. Appl. Crystallogr* **40:** 658–674

McDowall KJ, Lin-Chaol S & Cohen SN (1994) A + U content rather than a particular nucleotide order determines the specificity of RNase E cleavage. *J. Biol. Chem.* **269:** 10790–10796

McMullan G, Chen S, Henderson R & Faruqi AR (2009) Detective quantum efficiency of electron area detectors in electron microscopy. *Ultramicroscopy* **109:** 1126–43

Meyerson JR, Rao P, Kumar J, Chittori S, Banerjee S, Pierson J, Mayer ML & Subramaniam S (2014) Self-assembled monolayers improve protein distribution on holey carbon cryo-EM supports. *Sci. Rep.* **4:** 7084

Mikulecky PJ, Kaw MK, Brescia CC, Takach JC, Sledjeski DD & Feig AL (2004) Escherichia coli Hfq has distinct interaction surfaces for DsrA, rpoS and poly(A) RNAs. *Nat. Struct. Mol. Biol* **11:** 1206–1214

Millard P, Smallbone K & Mendes P (2017) Metabolic regulation is sufficient for global and robust coordination of glucose uptake, catabolism, energy production and growth in Escherichia coli. *PLOS Comput. Biol.* **13:** e1005396

Moffitt JR, Pandey S, Boettiger AN, Wang S & Zhuang X (2016) Spatial organization shapes the turnover of a bacterial transcriptome. *eLife* **5.**

Moll I (2003) Coincident Hfq binding and RNase E cleavage sites on mRNA and small regulatory RNAs. *RNA* **9:** 1308–1314

Moll JM, Sobti M & Mabbutt BC (2011) The Lsm Proteins: Ring Architectures for RNA Capture. In *RNA Processing* InTech Available at: http://www.intechopen.com/books/rna-processing/the-lsm-proteins-ring-architectures-for-rna-capture [Accessed July 3, 2019]

Møller T, Franch T, Højrup P, Keene DR, Bächinger HP, Brennan RG & Valentin-Hansen P (2002) Hfq. *Mol. Cell* **9:** 23–30

Moores A, Riesco AB, Schwenk S & Arnvig KB (2017) Expression, maturation and turnover of DrrS, an unusually stable, DosR regulated small RNA in Mycobacterium tuberculosis. *PloS one* **12:** e0174079

Morfeldt E, Taylor D, von Gabain A & Arvidson S (1995) Activation of alpha-toxin translation in Staphylococcus aureus by the trans-encoded antisense RNA, RNAIII. *EMBO J.* **14:** 4569–77

Morita T & Aiba H (2011) RNase E action at a distance: degradation of target mRNAs mediated by an Hfq-binding small RNA in bacteria. *Genes Dev.* **25:** 294–8

Morita T, Kawamoto H, Mizota T, Inada T & Aiba H (2004) Enolase in the RNA degradosome plays a crucial role in the rapid decay of glucose transporter mRNA in the response to phosphosugar stress in Escherichia coli. *Mol. Microbiol.* **54:** 1063–1075

Motlagh HN, Wrabl JO, Li J & Hilser VJ (2014) The ensemble nature of allostery. *Nature* **508:** 331–339

Murashko O & Lin-Chao S (2017) Escherichia coliresponds to environmental changes using enolasic degradosomes and stabilized DicF sRNA to alter cellular morphology. *Proc. Nat.l Acad. Sci.* **114:** E8025-E8034

Murashko ON, Kaberdin VR & Lin-Chao S (2012) Membrane binding of Escherichia coli RNase E catalytic domain stabilizes protein structure and increases RNA substrate affinity. *Proc. Natl. Acad. Sci. U. S. A.* **109:** 7019–7024

Murina V, Lekontseva N & Nikulin A (2013) Hfq binds ribonucleotides in three different RNA-binding sites. *Acta Crystallogr. D* **69:** 1504–1513

Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F & Vagin AA (2011) *REFMAC* 5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67:** 355–367

Nurmohamed S, McKay AR, Robinson C V. & Luisi BF (2010) Molecular recognition between Escherichia coli enolase and ribonuclease E. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66:** 1036–1040

Nurmohamed S, Vaidialingam B, Callaghan AJ & Luisi BF (2009) Crystal Structure of Escherichia coli Polynucleotide Phosphorylase Core Bound to RNase E, RNA and Manganese: Implications for Catalytic Mechanism and RNA Degradosome Assembly. *J. Mol. Biol.* **389:** 17–33

Ochoa De Alda JAG, Esteban R, Diago ML & Houmard J (2014) The plastid ancestor originated among one of the major cyanobacterial lineages. *Nat. Commun.* **5:** 4937

Oglesby-Sherrouse AG & Murphy ER (2013) Iron-responsive bacterial small RNAs: variations on a theme. *Metallomics* **5:** 276–86

Olsen AS, Møller-Jensen J, Brennan RG & Valentin-Hansen P (2010) C-Terminally Truncated Derivatives of Escherichia coli Hfq Are Proficient in Riboregulation. *J. Mol. Biol.* **404:** 173–182

Opdyke JA, Kang J-G & Storz G (2004) GadY, a small-RNA regulator of acid response genes in Escherichia coli. *J. Bacteriol* **186:** 6698–705

Orans J, Kovach AR, Hoff KE & Brennan RG Crystal structure of Escherichia coli Hfq DNA complex reveals multifunctional nucleic acid binding site. *TO BE PUBLISHED* Available at: https://www.rcsb.org/structure/5UK7

Otaka H, Ishikawa H, Morita T & Aiba H (2011) PolyU tail of rho-independent terminator of bacterial small RNAs is essential for Hfq action. *Prot. Nat. Acad. Sci.* **108:** 13059–13064

Panja S, Schu DJ & Woodson SA (2013) Conserved arginines on the rim of Hfq catalyze base pair formation and exchange. *Nucleic Acids Res.* **41:** 7536-46.

Panja S & Woodson SA (2012) Hfq proximity and orientation controls RNA annealing. *Nucleic Acids Res.* **40:** 8690–8697

Papenfort K, Bouvier M, Mika F, Sharma CM & Vogel J (2010) Evidence for an autonomous 5' target recognition domain in an Hfq-associated small RNA. *Prot. Nat. Acad. Sci.* **107:** 20435–20440

Papenfort K, Pfeiffer V, Lucchini S, Sonawane A, Hinton JCD & Vogel J (2008) Systematic deletion of Salmonella small RNA genes identifies CyaR, a conserved CRP-dependent riboregulator of OmpX synthesis. *Mol. Microbiol.* **68:** 890–906

Papenfort K, Sun Y, Miyakoshi M, Vanderpool CKK & Vogel J (2013) Small RNA-mediated activation of sugar phosphatase mRNA regulates glucose homeostasis. *Cell* **153:** 426-37.

Pause A & Sonenberg N (1992) Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. *EMBO J.* **11:** 2643–54

Peng Y, Curtis JE, Fang X & Woodson SA (2014) Structural model of an mRNA in complex with the bacterial chaperone Hfq. *Prot. Nat. Acad. Sci.* **111:** 17134–17139

Pereira J & Lupas AN (2018) The ancestral KH peptide at the root of a domain family with three different folds. *Bioinformatics* **34:** 3961–3965

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC & Ferrin TE (2004) UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25:** 1605–1612

Pfeiffer V, Papenfort K, Lucchini S, Hinton JCD & Vogel J (2009) Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. *Nat. Struct. Mol. Biol* **16:** 840–846

Plumbridge J, Bossi L, Oberto J, Wade JT & Figueroa-Bossi N (2014) Interplay of transcriptional and small RNA-dependent control mechanisms regulates chitosugar uptake in *Escherichia coli* and *Salmonella*. *Mol. Microbiol.* **92:** 648–658

Potter SC, Luciani A, Eddy SR, Park Y, Lopez R & Finn RD (2018) HMMER web server: 2018 update. *Nucleic Acids Res.* **46:** W200–W204

Prévost K, Salvail H, Desnoyers G, Jacques J-F, Phaneuf É & Massé E (2007) The small RNA RyhB activates the translation of shiA mRNA encoding a permease of shikimate, a compound involved in siderophore synthesis. *Mol. Microbiol.* **64:** 1260–1273

Prud'homme-Généreux A, Beran RK, Iost I, Ramey CS, Mackie GA & Simons RW (2004) Physical and functional interactions among RNase E, polynucleotide phosphorylase and the cold-shock protein, CsdA: Evidence for a 'cold shock degradosome'. *Mol. Microbiol.* **54:** 1409–1421

Purusharth RI, Klein F, Sulthana S, Jäger S, Jagannadham MV, Evguenieva-Hackenberg E, Ray MK & Klug G (2005) Exoribonuclease R interacts with endoribonuclease E and an RNA helicase in the psychrotrophic bacterium Pseudomonas syringae Lz4W. *J. Biol. Chem.* **280:** 14572–14578

Py B, Higgins CF, Krisch HM & Carpousis AJ (1996) A DEAD-box RNA helicase in the Escherichia coli RNA degradosome. *Nature* **381:** 169–172

Redko Y, Tock MR, Adams CJ, Kaberdin VR, Grasby JA & McDowall KJ (2003) Determination of the Catalytic Parameters of the N-terminal Half of Escherichia coli Ribonuclease E and the Identification of Critical Functional Groups in RNA Substrates. *J. Biol. Chem.* **278:** 44001–44008

Repoila F, Majdalani N & Gottesman S (2003) Small non-coding RNAs, co-ordinators of adaptation processes in Escherichia coli: The RpoS paradigm. *Mol. Microbiol.* **48:** 855–861

Rice JB & Vanderpool CK (2011) The small RNA SgrS controls sugar-phosphate accumulation by regulating multiple PTS genes. *Nucleic Acids Res.* **39:** 3806–19

Robinson KE, Orans J, Kovach AR, Link TM & Brennan RG (2014) Mapping Hfq-RNA interaction surfaces using tryptophan fluorescence quenching. *Nucleic Acids Res.* **42:** 2736–2749

Rohou A & Grigorieff N (2015) CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology* **192:** 216–221

Rust M, Bates M & Zhuang X (2006) Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods* **3:** 793-796

Santiago-Frangos A, Fröhlich KS, Jeliazkov JR, Ma\lecka EM, Marino G, Gray JJ, Luisi BF, Woodson SA & Hardwick SW (2019) Caulobacter crescentus Hfq structure reveals a conserved mechanism of RNA annealing regulation. *Prot. Nat. Acad. Sci. of the United States of America* **116:** 10978–10987

Santiago-Frangos A & Woodson SA (2018) Hfq chaperone brings speed dating to bacterial sRNA. *Wiley interdisciplinary reviews. RNA* **9:** e1475

Santos-Zavaleta A, Salgado H, Gama-Castro S, Sánchez-Pérez M, Gómez-Romero L, Ledezma-Tejeida D, García-Sotelo JS, Alquicira-Hernández K, Muñiz-Rascado LJ, Peña-Loredo P, Ishida-Gutiérrez C, Velázquez-Ramírez DA, Del Moral-Chávez V, Bonavides-Martínez C, Méndez-Cruz C-F, Galagan J & Collado-Vides J (2019) RegulonDB v 10.5: tackling challenges to unify classic and high throughput nowledge of gene regulation in E. coli K-12. *Nucleic Acids Res.* **47:** D212

Saramago M, Bárria C, dos Santos RF, Silva IJ, Pobre V, Domingues S, Andrade JM, Viegas SC & Arraiano CM (2014) The role of RNases in the regulation of small RNAs. *Curr. Opin. Microbiol.* **18:** 105–115

Sauer E, Schmidt S & Weichenrieder O (2012) Small RNA binding to the lateral surface of Hfq hexamers and structural rearrangements upon mRNA target recognition. *Prot. Nat. Acad. Sci.* **109:** 9396–9401

Sauer E & Weichenrieder O (2011a) Structural basis for RNA 3'-end recognition by Hfq. *Prot. Nat. Acad. Sci. of the United States of America* **108:** 13065–70

Sauer E & Weichenrieder O (2011b) Structural basis for RNA 3′-end recognition by Hfq. *Prot. Nat. Acad. Sci.* **108:** 13065–13070

Sauter C, Basquin J & Suck D (2003) Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from Escherichia coli. *Nucleic Acids Res.* **31:** 4091–4098

Schein A, Sheffy-Levin S, Glaser F & Schuster G (2008) The RNase E/G-type endoribonuclease of higher plants is located in the chloroplast and cleaves RNA similarly to the E. coli enzyme. *Rna* **14:** 1057–1068

Scheres SHW (2012) RELION: implementation of a Bayesian approach to cryo-EM structure determination. *Journal of structural biology* **180:** 519–30

Scheres S (2015) RELION. https://sbgrid.org/news/sbgrid-webinar-series

Schu DJ, Zhang A, Gottesman S & Storz G (2015) Alternative Hfq-sRNA interaction modes dictate alternative mRNA recognition. *EMBO J.* **34:** 2557-73

Schumacher MA (2002) Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. *EMBO J.* **21:** 3546–3556

Schümperli D & Pillai RS (2004) The special Sm core structure of the U7 snRNP: Far-reaching significance of a small nuclear ribonucleoprotein. *Cell. Mol. Life Sci.* **61:** 2560–2570

Shao Y, Feng L, Rutherford ST, Papenfort K & Bassler BL (2013) Functional determinants of the quorum-sensing non-coding RNAs and their roles in target regulation. *EMBO J.* **32:** 2158–2171

Sittka A, Lucchini S, Papenfort K, Sharma CM, Rolle K, Binnewies TT, Hinton JCD & Vogel J (2008) Deep Sequencing Analysis of Small Noncoding RNA and mRNA Targets of the Global Post-Transcriptional Regulator, Hfq. *PLoS Genet* **4:** e1000163

Sittka A, Pfeiffer V, Tedin K & Vogel J (2007) The RNA chaperone Hfq is essential for the virulence of Salmonella typhimurium. *Mol. Microbiol.* **63:** 193–217

Sittka A, Sharma CM, Rolle K & Vogel J (2009) Deep sequencing of Salmonella RNA associated with heterologous Hfq proteins in vivo reveals small RNAs as a major target class and identifies RNA processing phenotypes. *RNA Biol* **6:** 266–275

Soper T, Mandin P, Majdalani N, Gottesman S & Woodson SA (2010) Positive regulation by small RNAs and the role of Hfq. *Prot. Nat. Acad. Sci.* **107:** 9602–7

Soper TJ & Woodson SA (2008) The rpoS mRNA leader recruits Hfq to facilitate annealing with DsrA sRNA. *RNA* **14:** 1907–17

Soupene E, van Heeswijk W, Plumbridge J, Stewart V, Bertenthal D, Lee H, Prasad G, Paliy O, Charernnoppakul P & Kustu S (2003) Physiological Studies of Escherichia coli Strain MG1655: Growth Defects and Apparent Cross-Regulation of Gene Expression. *J. Bacteriol.* **185:** 5611–5626

Spence JCH (2003) High-Resolution Electron Microscopy 3rd ed. Oxford University Press

Storz G, Opdyke JA & Zhang A (2004) Controlling mRNA stability and translation with small, noncoding RNAs. *Curr. Opin. Microbiol.* **7:** 140–144

Storz G, Vogel J & Wassarman KM (2011) Regulation by Small RNAs in Bacteria: Expanding Frontiers. *Mol. Cell* **43:** 880–891

Strahl H, Turlan C, Khalid S, Bond PJ, Kebalo JM, Peyron P, Poljak L, Bouvier M, Hamoen L, Luisi BF & Carpousis AJ (2015) Membrane Recognition and Dynamics of the RNA Degradosome. *PLoS Genet.* **11:** 1–23

Sulthana S, Basturea GN & Deutscher MP (2016) Elucidation of pathways of ribosomal RNA degradation: an essential role for RNase E. *RNA* **22:** 1163–71

Suzuki K, Shimizu M, Sasaki N, Ogawa C, Minami H, Sugimoto H & Watanabe T (2016) Regulation of the chitin degradation and utilization system by the ChiX small RNA in Serratia marcescens 2170. *Biosci. Biotechnol. Biochem.* **80:** 376–385

Szczesny RJ, Borowski LS, Malecki M, Wojcik MA, Stepien PP & Golik P (2012) RNA Degradation in Yeast and Human Mitochondria. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1819:** 1027–1034

Tan EM & Kunkel HG (1966) Characteristics of a soluble nuclear antigen precipitating with sera of patients with systemic lupus erythematosus. *J. Immunol.* **96:** 464–71

Tanner NK, Cordin O, Banroques J, Doère M & Linder P (2003) The Q motif: A newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. *Mol. Cell* **11:** 127–138

Tegunov D & Cramer P (2018) Real-time cryo-EM data pre-processing with Warp. *bioRxiv***:** 338558

Teplova M, Malinina L, Darnell J, Song J, Lu M, Abagyan R, Musunuru K, Teplov A, Burley S, Darnell R & Patel D (2011) Protein-RNA and Protein-Protein Recognition by Dual KH1/2 Domains of the Neuronal Splicing Factor Nova-1. *Structure* **19:** 930-944

Tompa P & Csermely P (2004) The role of structural disorder in the function of RNA and protein chaperones. *The FASEB Journal* **18:** 1169–1175

Torrisi A (2011) Nanoworld in 3-D. *Opticon1826* **10**: Commentaries and Research Notes, item 2

Törö I, Thore S, Mayer C, Basquin J, Séraphin B & Suck D (2001) RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex. *EMBO J.* **20:** 2293–2303

Tree JJ, Granneman S, McAteer SP, Tollervey D & Gally DL (2014) Identification of Bacteriophage-Encoded Anti-sRNAs in Pathogenic Escherichia coli. *Mol. Cell* **55:** 199–213

Updegrove TB, Zhang A & Storz G (2016) Hfq: the flexible RNA matchmaker. *Curr. Opin. Microbiol.* **30:** 133–138

Urban JH & Vogel J (2007) Translational control and target recognition by Escherichia coli small RNAs in vivo. *Nucleic Acids Res.* **35:** 1018–1037

Valverde R, Edwards L & Regan L (2008) Structure and function of KH domains. *FEBS* **275:** 2712–2726

Valverde R, Pozdnyakova I, Kajander T, Venkatraman J & Regan L (2007) Fragile X Mental Retardation Syndrome: Structure of the KH1-KH2 Domains of Fragile X Mental Retardation Protein. *Structure* **15:** 1090–1098

Vanderpool CK & Gottesman S (2007) The novel transcription factor SgrR coordinates the response to glucose-phosphate stress. *J. Bacteriol* **189:** 2238–48

Večerek B, Rajkowitsch L, Sonnleitner E, Schroeder R & Bläsi U (2008) The C-terminal domain of Escherichia coli Hfq is required for regulation. *Nucleic Acids Res.* **36:** 133–143

Viegas SC & Arraiano CM (2008) Regulating the regulators: How ribonucleases dictate the rules in the control of small non-coding RNAs. *RNA Biology* **5:** 230–243

Vincent HA, Henderson CA, Ragan TJ, Garza-Garcia A, Cary PD, Gowers DM, Malfois M, Driscoll PC, Sobott F & Callaghan AJ (2012) Characterization of Vibrio cholerae Hfq Provides Novel Insights into the Role of the Hfq C-Terminal Region. *J. Mol. Biol.* **420:** 56–69

Vitrobot_500x500 (2019) Available at: https://www.fei.com/uploadedImages/FEISite/Pages/Products/Specialty_Products/Falcon(1)/Models/Vitrobot_500x500.png?n=1650 [Accessed December 16, 2019]

Vogel J & Luisi BF (2011) Hfq and its constellation of RNA. *Nat. Rev. Microbiol.* **9:** 578–589

Vogel J & Wagner EGH (2007) Target identification of small noncoding RNAs in bacteria. *Curr. Opin. Microbiol.* **10:** 262–270

Wagner EGH (2013) Cycling of RNAs on Hfq. *RNA Biology* **10:** 619–626

Wagner EGH & Romby P (2015) Small RNAs in Bacteria and Archaea: Who They Are, What They Do, and How They Do It. *Adv. Genet.* **90:** 133–208

Wang J, Rennie W, Liu C, Carmack CS, Prévost K, Caron M-P, Massé E, Ding Y & Wade JT (2015) Identification of bacterial sRNA regulatory targets using ribosome profiling. *Nucleic Acids Res.* **43:** 10308–20

Wang W, Wang L, Wu J, Gong Q & Shi Y (2013) Hfq-bridged ternary complex is important for translation activation of rpoS by DsrA. *Nucleic Acids Res.* **41:** 5938–5948

Wang W, Wang L, Zou Y, Zhang J, Gong Q, Wu J & Shi Y (2011) Cooperation of Escherichia coli Hfq hexamers in DsrA binding. *Genes Dev.* **25:** 2106-17

Wang Z & Wu M (2015) An integrated phylogenomic approach toward pinpointing the origin of mitochondria. *Sci. Rep.* **5:** 7949

Weston A & Sommerville J (2006) Xp54 and related (DDX6-like) RNA helicases: Roles in messenger RNP assembly, translation regulation and RNA degradation. *Nucleic Acids Res.* **34:** 3082–3094

Wilson N, Pandey P, Beanland R, Young R, Kinloch I, Gong L, Liu Z, Suenaga K, Rourke J, York S & Sloan J (2009) Graphene Oxide: Structural Analysis and Application as a Highly Transparent Support for Electron Microscopy. *ACS Nano* **3**: 2547-2556

Wilusz CJ & Wilusz J (2013) Lsm proteins and Hfq. *RNA Biology* **10:** 592–601

Woodson SA, Panja S & Santiago-Frangos A (2018) Proteins That Chaperone RNA Regulation. *Microbiol. Spectr.* **6:** RWR-0026-2018

Worrall J, Górna M, Crump N, Phillips L, Tuck A, Price A, Bavro V & Luisi B (2008) Reconstitution and Analysis of the Multienzyme Escherichia coli RNA Degradosome. *J. Mol. Biol* **382:** 870-883

Xing Z, Ma WK & Tran EJ (2019) The DDX5/Dbp2 subfamily of DEAD-box RNA helicases. *Wiley Interdiscip. Rev. RNA* **10:** e1519

Xiol J, Spinelli P, Laussmann MA, Homolka D, Yang Z, Cora E, Couté Y, Conn S, Kadlec J, Sachidanandam R, Kaksonen M, Cusack S, Ephrussi A & Pillai RS (2014) RNA Clamping by Vasa Assembles a piRNA Amplifier Complex on Transposon Transcripts. *Cell* **157:** 1698–1711

Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG & Court DL (2000) An efficient recombination system for chromosome engineering in Escherichia coli. *Prot. Nat. Acad. Sci.* **97:** 5978–83

Zhang A, Schu DJ, Tjaden BC, Storz G & Gottesman S (2013) Mutations in Interaction Surfaces Differentially Impact E. coli Hfq Association with Small RNAs and Their mRNA Targets. *J. Mol. Biol.* **425:** 3678–3697

Zhang A, Wassarman KM, Ortega J, Steven AC & Storz G (2002) The Sm-like Hfq Protein Increases OxyS RNA Interaction with Target mRNAs. *Mol. Cell* **9:** 11–22

Zhang A, Wassarman KM, Rosenow C, Tjaden BC, Storz G & Gottesman S (2003) Global analysis of small RNA and mRNA targets of Hfq. *Mol. Microbiol.* **50:** 1111–1124

Zhang K (2016) Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193:** 1–12

Zhao B & Houry WA (2010) Acid stress response in enteropathogenic gammaproteobacteria: an aptitude for survival. *Biochemistry and Cell Biology* **88:** 301–314

Zheng SQ, Palovcak E, Armache J-P, Verba KA, Cheng Y & Agard DA (2017) MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods.* **14:** 331–332

Zhou L, Zhou Y, Hang J, Wan R, Lu G, Yan C & Shi Y (2014) Crystal structure and biochemical analysis of the heptameric Lsm1-7 complex. *Cell Res.* **24:** 497–500